Discovery of novel bioactive compounds: Molecular awakening of silent clusters in *Streptomyces natalensis*

by

Rui Filipe Ferreira Vilar da Silva

bio09084@fe.up.pt

Integrated Master in Bioengineering – Molecular Biotechnology

Master Thesis

September 2014

Discovery of novel bioactive compounds: Molecular awakening of silent clusters in *Streptomyces natalensis*

by

Rui Filipe Ferreira Vilar da Silva

Submitted in partial fulfilment of the requirements for the
Master of Bioengineering degree in the branch of Molecular Biotechnology at the Faculty of Engineering and Institute of Biomedical Sciences Abel Salazar of the University of Porto

Abstract

The healthcare system is facing a serious problem: the emergence of antimicrobial drug resistance. However, antimicrobial drug discovery and development has decelerated considerably throughout the past decades, creating an urgent need for the development of new strategies for antibiotic discovery. *Streptomyces* are the native producers of approximately 75% of the antibacterial drugs in clinical use and it is believed that we are only taking advantage of a small part of their production potential.

In this work, we performed an *in silico* genome mining approach combined with genetic engineering, for the development of novel hit compounds that will ultimately feed the drug discovery pipeline.

Genome mining analysis of *S. natalensis* ATCC 27448 unveiled 29 biosynthetic gene clusters coding for unknown specialized products. Six of these were submitted for an *in silico* and transcription analysis. Three predicted positive cluster situated regulators (CSRs), located within three different clusters and that showed no expression under laboratory conditions were selected for overexpression. RT-PCR analysis of *S. natalensis* strains overexpressing the CSR encoding genes *SNA01173*, *SNA01239* and *SNA06246* did not show up-regulation of structural genes at the analysed time point. Comparative metabolic profiling assays by bioassays and HPLC revealed different production patterns which suggest that novel biologically active compounds might being produced. Such products may further feed the drug discovery pipeline. Finally, overexpression of *SNA01239* and *SNA01173* also significantly altered pimaricin yield.

Supervisor: Marta Vaz Mendes, PhD

Co-supervisor: Tiago Beites, PhD

Affiliation: IBMC – Institute for Molecular and Cell Biology

You can't put a limit on anything. The more you dream, the farther you get.

Michael Phelps

Acknowledgments

I would like to thank my supervisor, Doctor Marta Vaz Mendes, for the opportunity she gave me to work, not only at a group that I am keen on, but also in a very interesting, ambitious and well planned project. I am also grateful for her guidance and help during this semester. Since I have been working at the IBMC – *Cell and Applied Microbiology* group I feel that, besides improving my technical capabilities in regard of scientific experiments, I have developed my autonomy and improved my cognitive and research capabilities. This fact would not be possible without some good and committed leadership. For all of that I am truly thankful.

I also would like to thank my co-supervisor, Tiago, for being there every day, teaching me new experiments, encouraging me when I most needed it and listening to all my doubts, fears and questions. His patience with me was admirable and priceless.

My thanks to IBMC – *Cell and Applied Microbiology* group members Rita Vilaça, Rute Oliveira, Sílvia Pires and Nuno Carvalho for their interest, availability and encouragement. They had a boundless patience, answering all my never-ending questions at any time. Moreover, altogether, they have created the most amazing atmosphere at the laboratory, making daily life richer and more fun. Because of that, I am tremendously grateful to all.

I am also thankful to the members of IBMC - *Redox Cell Signalling* and *Bioengineering and Synthetic Microbiology* groups for all the help and friendship.

Special thanks to my very good friends: João Pedro Almeida, Sérgio Baptista, Liliana Pinho, Henrique Duarte, Frederico Carpinteiro, Miguel Ferraria, Sara Campos and Vitor Pereira, for their love, unconditional support and encouragement. Furthermore, I would also like to acknowledge my swimming coach, Rui Borges, as well as my swim mates Ana Neto, Bernardo Valente, Bruno Silva, João Travanca, Júlio Gonçalves, Maria Amorim, Nuno Guerra and Tomás Silva, for their friendship and willingness to put up with all my (monotonous) laboratory conversations.

To Connor Cassidy, Máté Ambrus, Schuyler Ellis, Anis Ouguenoune, André Meireles, Miren Otaegi, Tímea Balogh, Ágnes Duzs, Hegedűs Botond, Carlos Martínez, Laura Felix, despite the fact that we do not talk that often, mostly due to the geographical distance between us, I address my sincerest thanks. They are all very important people to me and I have learned a lot from each one of them.

Last, but not the least, a few words to my family: mum, dad, Mitó, Guigo, Rozinho, Ná, Zézé, Vovó Dilita, Inha and Nana! Since I remember I have always been taught to study and work as much as I can so one day I may have the reward. Furthermore, they have also been

there for me, at any time, no matter what. That are no words that can describe how truly thankful I am to all of them. Their love and help will never be forgotten.

I have saved a last couple of words to my grandfather, Zé. He was an inspiration of what a human being must be: hard worker, but, simultaneously, generous person and lover of life. Wherever you are, “Obrigado”!

Thank you all for making this thesis possible and for making me who I am today.

List of contents

List of figures

List of tables

List of abbreviations

1-deoxy-D-xylulose 5-phosphate	DOX
1-deoxy-D-xylulose 5-phosphate synthase	DXS
2-methylisoborneol	MIB
Acyl carrier protein	ACP
Acyltransferase	AT
Adenylation	A
Amino acids	aa
Antimicrobial resistance	AMR
Aromatase	ARO
Base pairs	bp
cAMP receptor protein	CRP
Chain length factor	CLF
Chalcone synthase	CHS
Cluster-situated regulators	CSR
Coenzyme A	CoA
Condensation	C
Cyclase	CYC
Dehydratase	DH
Digoxygenin	DIG
Dimethyl sulphoxide	DMSO
Dimethylallyl diphosphate	DMAPP
Enoylreductase	ER
Ethylenediamine tetracetic acid	EDTA
European Molecular Biology Laboratory	EMBL
Farnesyl diphosphate	FPP
Farnesyl diphosphate synthase	FPPS
Fumarate and nitrate reduction regulatory protein	FNR
Geranyl diphosphate	GPP
Geranyl diphosphate synthase	GPPS
Geranylgeranyl diphosphate	GGPP
Geranylgeranyl diphosphate synthase	GGPPS
Glyceraldehyde-3-phosphate	G3P
Helix-turn-helix	HTH
Hydroxylase	HYDR
Isopentenyl diphosphate	IPP
Isopentenyl diphosphate isomerase	IPPI
Isopropyl β-D-1-thiogalactopyranoside	IPTG
Ketosynthase	KS
Luria-Bertani	LB
Major facilitator superfamily	MFS
Mass Spectrometer	MS
Methylerythritol phosphate	MEP
Mevalonate	MVA
Monooxygenase	MO
Multidrug resistance	MDR
National Centre for Biotechnology Information	NCBI
Nonribossomal peptide	NRPS

Nuclear magnetic resonance	NMR
O-methyltransferase	O-MTR
Optical density	OD
Over-night	O/N
Peptidyl carrier protein	PCP
Phytoene	PE
Phytoene dehydrogenase	PD
Phytoene synthase	PSY
Polyketide synthase	PKS
Polymerase chain reaction	PCR
ppGpp 5'-diphosphate 3'-diphosphate	ppGpp
quantitative RT-PCR	RT-qPCR
Retention time	RT
Reverse transcriptase polymerase chain reaction	RT-PCR
Ribonucleic acid	RNA
RNA polymerase	RNAP
Room temperature	RT
Rotations per minute	rpm
Squalene synthase	SQS
Thioesterases	TE
Tripeptidyl aminopeptidase	TAP
Volume per volume	v/v
Weight per volume	w/v
World Health Organization	WHO
β-ketoreductase	KR

1. Introduction

1.1 The need for novel compounds

Antibiotics are agents that, due to their ability to kill microorganisms or suppressing their growth, have been widely used for the treatment of infectious diseases. However, the improper or unsuitable use of these drugs together with a poor prevention and control of infectious diseases, gives rise to the phenomenon known as antimicrobial resistance (AMR).

Infections caused by resistant microorganisms often fail to respond to conventional treatment, resulting in prolonged illness, greater risk of death and higher costs. Each year in the European Union more than 25 000 people die as a direct result of the antibiotic resistant bacteria and is estimated economical loses of 1.5 billion euros (European Centre for Disease Prevention and Control, 2013). In fact, according to the World Health Organization (WHO), the world is heading towards a post-antibiotic era in which common infections and minor injuries, treatable for decades, can kill again.

It should be noted that AMR is a not a novel problem. It has been present in society over the last decades and continues to increase. Moreover, multi-drug resistant Gram-negative bacteria such as *Escherichia coli*, *Klebsiella pneumonia*, *Acinetobacter baumannii* and *Pseudomonas aeruginosa* have been red-flagged as antimicrobial priority targets (*European Centre for Disease Prevention and Control*, 2009, 2013).

Therefore, there is a political and social demand to find a solution for antimicrobial resistance. Besides imposing preventive measures, such as the prudent use of antibiotics, the development of new antimicrobials to control antibiotic-resistant bacteria is needed.

Antimicrobial drug discovery and development has slowed down significantly during the last decades (Watve *et al.*, 2001) (Fig. 1). Only a handful of antibiotics were developed and have hit the market in recent decades. The low profitability of antibiotics' market from the point of view of pharmaceutical industry (these drugs are only taken for a short time and only when a patient suffers specific infections) has contributed for the lack of investment on the development of new antibiotics. However the need to fight AMR sets antibacterial drugs as a market of billions of dollars (Devasahayam *et al.*, 2010). Moreover, the FDA and its European counterpart, EMA, are now actively committed in encouraging the pharmaceutical industry on developing new antimicrobials and in streamlining the validation of those drugs (*Food and Drug Administration*, 2011).

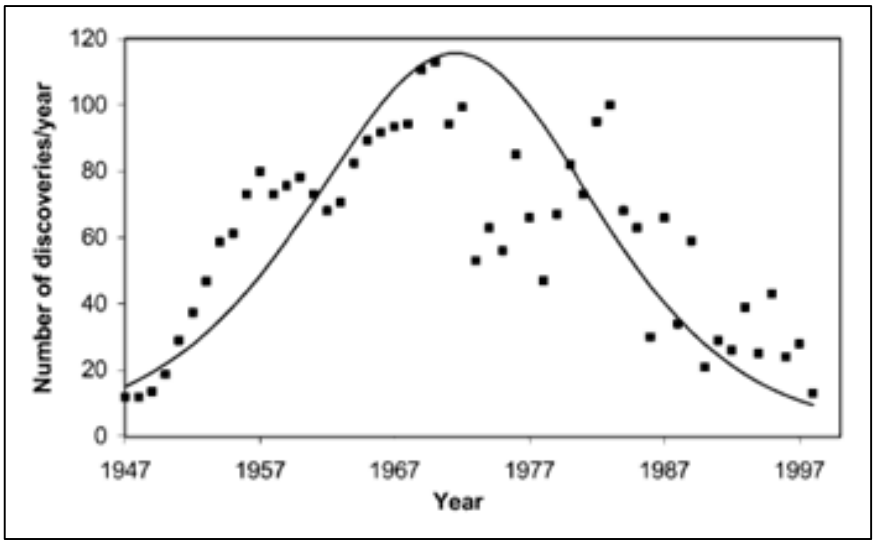

Figure 1 – Logistic model reflecting the trend of antimicrobial compounds discoveries per year, from 1947 to 1997 [Adapted from Watve, *et al.*, 2001].

1.2 *Streptomyces* and specialized metabolites

Streptomyces, the largest genus of the Actinobacteria phylum, are Gram-positive, filamentous, obligated aerobes, soil-dwelling bacteria that present a high G+C content in the genome (Chater and Chandra, 2006). Their life cycle begins with the germination of a spore that leads to the formation of the vegetative mycelium which, after a morphological differentiation, gives rise to the emergence of aerial hyphae (Hardisson *et al.*, 1978). New spores are originated once the aerial hyphae are divided into unigenomic compartments by the septation of the aerial mycelium (Fig. 2).

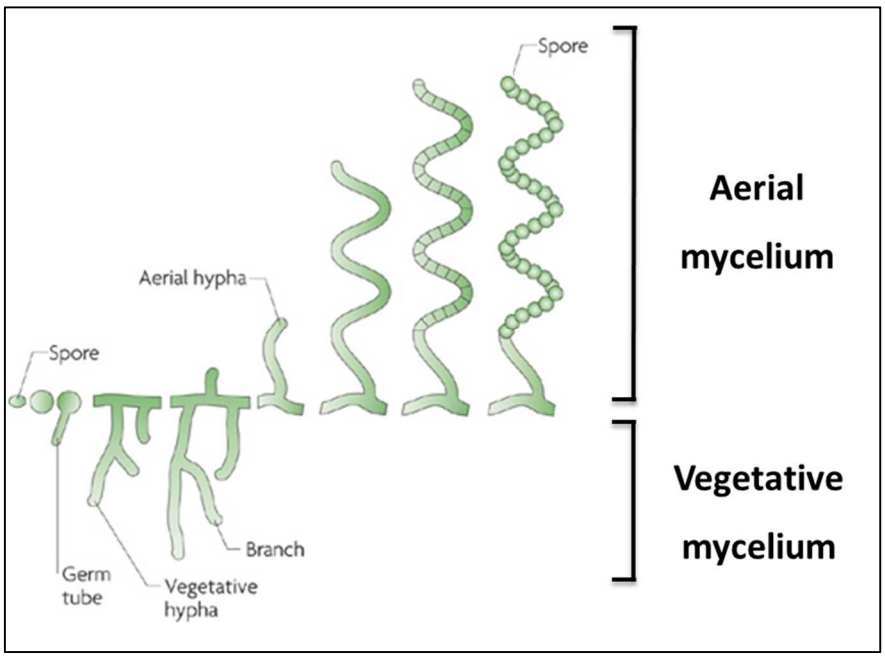

Figure 2 – The life cycle of *Streptomyces* spp. [Adapted from Flardh and Buttner, 2009].

Over the years, *Streptomyces* bacteria have attracted the interest of the scientific community due to their ability to produce a wide range of secondary metabolites (a.k.a. as specialized metabolites) of clinical, agricultural and biotechnological value. Nearly 75% of the antibacterial drugs in clinical use are, or are derived from, natural products synthesized by *Streptomyces* or other closely related actinomycetes species (Berdy, 2005). Additionally, *Streptomyces* also produce antifungal (amphotericin), immunosuppressive (e.g. rapamycin and FK506), neuroprotectant (e.g. meridamycin) and anticancer compounds (e.g. doxorubicin and epothilone), among others. About 7 000 secondary bioactive metabolites have already been isolated from *Streptomyces* cultures (Berdy, 2005). However, it is estimated that more than 100 000 new bioactive compounds are still waiting to be discovered, which underlines the tremendous unclaimed potential of these bacteria (Watve *et al.*, 2001). Altogether, these facts highlight *Streptomyces* as a very promising and powerful source of compounds that could be used to fight antibiotic resistance.

Secondary metabolites are considered non-essential for the development and survival of microorganisms but there is no doubt that they play an important ecological role in their ability to adapt to environmental changes. The biosynthesis of these compounds generally occurs in a growth-dependent manner and is regulated by environmental signals.

Current developments in DNA sequencing technologies have made genome sequencing faster, cheaper and more efficient. By October 2014 so far, 19 *Streptomyces* genomes have been completely sequenced and annotated and 158 more are available in scaffolds at the NCBI database (http://www.ncbi.nlm.nih.gov/genome/). *Streptomyces* spp. present large linear chromosomes (approximately 8–10 Mb) with over 20 putative secondary metabolite biosynthetic gene clusters per strain on average. The secondary metabolism diversity of these bacteria is reflected on the structural diversity of the proteins that govern the biosynthesis of polyketides, peptides, bacteriocins, terpenoids, aminoglycosides, and other natural products (Staunton and Weissman, 2001, Marahiel and Essen, 2009, Moore, 2008). Nonetheless, only a small fraction of them is expressed under laboratory controlled conditions. These biosynthetic clusters with reduced or no expression are known as silent or cryptic clusters and their activation is an auspicious strategy to produce novel bioactive compounds with the potential to revolutionize drug discovery pipelines.

1.3 Genome mining

Recent advances in DNA sequencing techniques have led to a relatively easy and inexpensive access to large numbers of biosynthetic gene clusters via *in silico* genome mining. The identification of the gene clusters is facilitated, firstly, by the fact that the resulting

enzymes can be categorised into a small set of families (e.g. terpene cyclases, NRPS, PKS) and, secondly, by the characteristic patterns shared by each protein family at the protein level, which is reflected at the genomic level. These two facts, together with the tendency of biosynthetic genes to be co-localized in the genome and organized in clusters, provide the basis for homology-driven genome mining, where conserved protein motifs and consensus sequences are used to identify the *loci* of putative biosynthetic clusters (Gross, 2007, Nett *et al.*, 2009).

Currently there are several available bioinformatics tools for the *in silico* analysis of specialized metabolite biosynthetic gene clusters. These softwares differ from each other on their usability, which, in turn, depends on the following factors: (1) the quality of the sequence data, the user's needs, and the user computer proficiency; (2) the capability of performing other functions such as predicting substrate specificity of polymerization enzymes or generating putative 3D structures; and (3) the ability of detecting entire secondary metabolites gene clusters in microbial genomes (Fedorova *et al.*, 2012). The only ones that can actually identify whole biosynthetic gene clusters, rather than individualized genes, as well as different types of secondary metabolites are the Secondary Metabolite Unknown Regions Finder (SMURF; http://jcvi.org/smurf/index.php) and the antibiotics & Secondary Metabolite Analysis Shell (antiSMASH; http://antismash.secondarymetabolites.org/) (Khaldi *et al.*, 2010, Medema *et al.*, 2011). Both tools incorporate and integrate multiple algorithms dedicated to more specific analysis, such as Cluster Sequence Analyzer (CLUSEAN), NRPSPredictor, Cluster Scanner (ClustScan), Structure Based Sequence Analysis of Polyketide Synthases (SBSPKS), or Natural Product searcher (NP.searcher).

Finally, it should also be mentioned that what distinguishes antiSMASH from SMURF is the fact that the latter is specific to analyse the fungal secondary metabolites gene clusters, while the first was preferentially developed to examine bacterial genomes. Furthermore, antiSMASH has some other advantages which have turned it a standard tool to analyse the bacterial potential to produce specialized metabolites. Those advantages are: (1) it is comprehensive, rapid and has a good usability; (2) it examines neighbouring regions on the chromosome to predict the entire gene clusters; (3) it allows running additional analyses such as clusters of orthologous groups, (4) it integrates into the pipeline other prediction tools, and (5) it combines more functionalities than any other similar programme (Medema *et al.*, 2011, Blin *et al.*, 2013).

1.4 Silent clusters activation strategies

Several different approaches for the activation of silent clusters in *Streptomyces* have already been described. Ribosome engineering is one of them. This method arose from the initial finding of the production of the blue-pigmented antibiotic actinorhodin by a *Streptomyces lividans* strain with a mutation on the *rpsL* gene, which encodes the 30S ribosomal subunit protein S12 (Shima *et al.*, 1996). Later, the bacterial alarmone ppGpp was found to bind to RNA polymerase (RNAP) (Artsimovitch *et al.*, 2004), leading to the production of antibiotics (Bibb, 2005, Ochi, 2007). This last discovery suggested that RNAP might be modified in order to mimic ppGpp-bound form and activate the expression of secondary metabolite biosynthetic gene clusters (Lai *et al.*, 2002, Xu *et al.*, 2002). Furthermore, studies with *Streptomyces mauvecolor* 631689, a strain with no antibacterial activity in any tested medium, demonstrated that when the strain was mutated on the *rpoB* gene, which encodes the RNAP β subunit, and/or on *rpsL,* becomes able to produce a family of antibiotics known as piperidamycins. The result was ascribed to the increased affinity of the mutant RNAP to the gene promoters (Hosaka *et al.*, 2009).

Another possible strategy to unlock the clusters expression is to grow bacteria under particular stress conditions or to vary culture conditions (Scherlach and Hertweck, 2009, Williams *et al.*, 2008, Bok *et al.*, 2009). However, this approach is largely empirical, as the physiological conditions responsible for the activation of the silent gene clusters are not clearly understood and it is often not possible to predict the complex regulatory circuits involved in pathway regulation (Scherlach and Hertweck, 2006).

Communication between microorganisms is also another way to induce the expression of silent biosynthesis gene clusters. Its potential was already demonstrated through the intimate interaction between the fungus *Aspergillus nidulans* and *Streptomyces hygroscopicus*, which resulted in the activation of a polyketide gene cluster (Schroeckh *et al.*, 2009), or by the co-culturing of *Streptomyces clavuligerus* with the methicillin-resistant *Staphylococcus aureus,* which gave rise to the activation of the holomycin biosynthesis by the *Streptomyces* strain (Charusanti *et al.*, 2012). However, likewise the previous strategy, the activation of the silent gene clusters is not completely understood and it is hard to predict what are the pathways involved.

One other approach to access the compounds uncovered behind silent clusters is to induce their expression in a surrogate strain (Gomez-Escribano and Bibb, 2011). This is a very useful approach while working with difficult-to-culture microorganisms, such as rare terrestrial

and marine actinomycetes (Fenical and Jensen, 2006, Baltz, 2008) or myxobacteria (Wenzel and Müller, 2009), especially at a large scale.

A potential candidate for heterologous protein production is the Gram-negative bacterium *Escherichia coli*, since it is well-known and its manipulation well established (Terpe, 2006). In fact, *E. coli* is the most commonly used organism for this kind of experiments. The expression of a cryptic terpene synthase gene of *Streptomyces coelicolor* in *E. coli* was already achieved and it led to the identification of a new sesquiterpene, epi-isozaene (Lin *et al.*, 2006). However, the use of a different bacterial genus as a host, such as *E. coli*, may result in difficulties, namely by the existence of rare codons or an unsuited metabolic background (Kane, 1995, Goldman *et al.*, 1995).

For this reason *Streptomyces avermitilis* and *S. coelicolor*, whose genome has already been sequenced, have been utilized for the development of expression hosts (Komatsu *et al.*, 2010, Gomez-Escribano and Bibb, 2012). While bellowing to the same bacterial genus, there is a higher chance for these bacteria to have the required supply of primary metabolic precursors for a successful production of the secondary metabolites, as well as reducing agents and energy source, also derived from primary metabolism, such as $NAD(P)^+$ and ATP (Komatsu *et al.*, 2010). Additionally, the mentioned drawbacks, such as the existence of rare codons, should no longer exist. In fact, it has already been shown that *S. avermitilis* is able to produce cephamycin C (from *S. clavuligerus* ATCC 27064) or pladienolide (from *S. platensis* Mer-11107) in higher levels than those of the native-producing species (Komatsu *et al.*, 2010). Nevertheless, despite the described advantages, some other technical challenges, such as the transferring of large gene clusters or reduced expression may also stand as an obstacle and compromise the efficacy of this approach. Therefore, new molecular tools addressing those issues, such as the construction of vectors for the shifting of large DNA fragments, are being developed (Jones *et al.*, 2013).

Along with heterologous expression, the recent developments on high-throughput DNA sequencing technologies also contributed to reinforce the cluster-situated regulators (CSRs) manipulation as a strategy to activate silent gene clusters. The first molecular analyses of secondary metabolite biosynthetic gene clusters highlighted the presence of regulatory genes, which often have major effects on the production levels of the related antibiotic (Liu *et al.*, 2013). These CSR regulators, formerly known as pathway-specific, are located within the biosynthetic gene cluster and its regulating activity is often limited to that biosynthetic cluster, although some studies have revealed that pathway-specific regulators can exert a wider effect on global transcription patterns (Huang *et al.*, 2005).

It is a fact that CSRs can directly control the expression and activity of biosynthetic gene clusters. However, it should still be noted that the action of pleiotropic regulators, presenting a wide-range effect over cell metabolism, may also result in the modulation of the biosynthesis of secondary metabolites (Gust *et al.*, 2004) (Fig. 3). For example, in *S. coelicolor*, the two-component system PhoR-PhoP, besides being the major signal transduction system for phosphate control, it has also been associated with the negative regulation of some secondary, and even primary metabolic pathways (Allenby *et al.*, 2012, Liu *et al.*, 2013). Once phosphorylated by PhoR, PhoP binds to the promoter region of the target gene, influencing its expression. The promoter of *afsS*, which encodes a global activator of antibiotic production, is a target, and its repression affects indirectly actinorhodin and undecylprodiginine production in *S. coelicolor* (Santos-Beneit *et al.*, 2009). The CSR encoding gene, *cdaR*, which regulates the calcium-dependent antibiotic biosynthesis in *S. coelicolor*, is regulated directly by PhoP (Allenby *et al.*, 2012). Furthermore, PhoR-PhoP system also plays an important role in the regulation of nitrogen metabolism in *S*. *coelicor*, due to PhoP ability to bind to the promoter region of *glnR*. GlnR is the main nitrogen regulator and controls the expression of several genes involved in nitrogen metabolism. In fact the nitrogen metabolism, through a different two-component regulatory system AfsQ1/Q2, may also stand as another example of an indirect effect over CSRs. AfsQ1 has been revealed as a direct repressor of primary nitrogen assimilation genes and as an activator of several CSRs such as *actII*-ORF4, *redZ*, and *cdaR* (Wang *et al.*, 2012). Moreover, AfsQ1 recognizes a moderately conserved pair of sequences similar to that recognized by the nitrogen regulator GlnR, leading to a competition between GlnR activation and AfsQ1 repression (Wang *et al.*, 2012).

Understanding the major role of CSRs highlights their genetic manipulation as a straightforward approach to activate silent gene clusters either by overexpressing activators or deleting repressors. In the year of 2007, Bergmann and co-workers demonstrated that the expression of a putative CSR encoding gene under the control of an inducible promoter causes the expression of a normally silent gene cluster in *Aspergillus nidulans* upon the addition of the inducer (Bergmann *et al.*, 2007). Aspyridones, the metabolic products of that gene cluster, were identified by comparative metabolic profiling of the wild-type and mutant strains and spectroscopic analyses showed them to have novel structures (Bergmann *et al.*, 2007). In the same way, the overexpression of the CSR *samR0484*, a LAL-family regulatory gene of *Streptomyces ambofaciens,* activated the expression of a cryptic type I PKS cluster, resulting in the production the stambomycins, macrolides with anti-proliferative activity against cancer lines (Laureti *et al.*, 2011). Concordantly with the previous experiments, the removal of *scbR2*, a gene encoding an inhibitory protein, allowed initiating the expression of the antibacterial

compound abCPK, from *S. coelicolor* (Gottelt *et al.*, 2010), and, thus, proving the utility of this approach for discovering new natural products of silent biosynthetic gene clusters that are not expressed in laboratory cultures (Scheffler *et al.*, 2013).

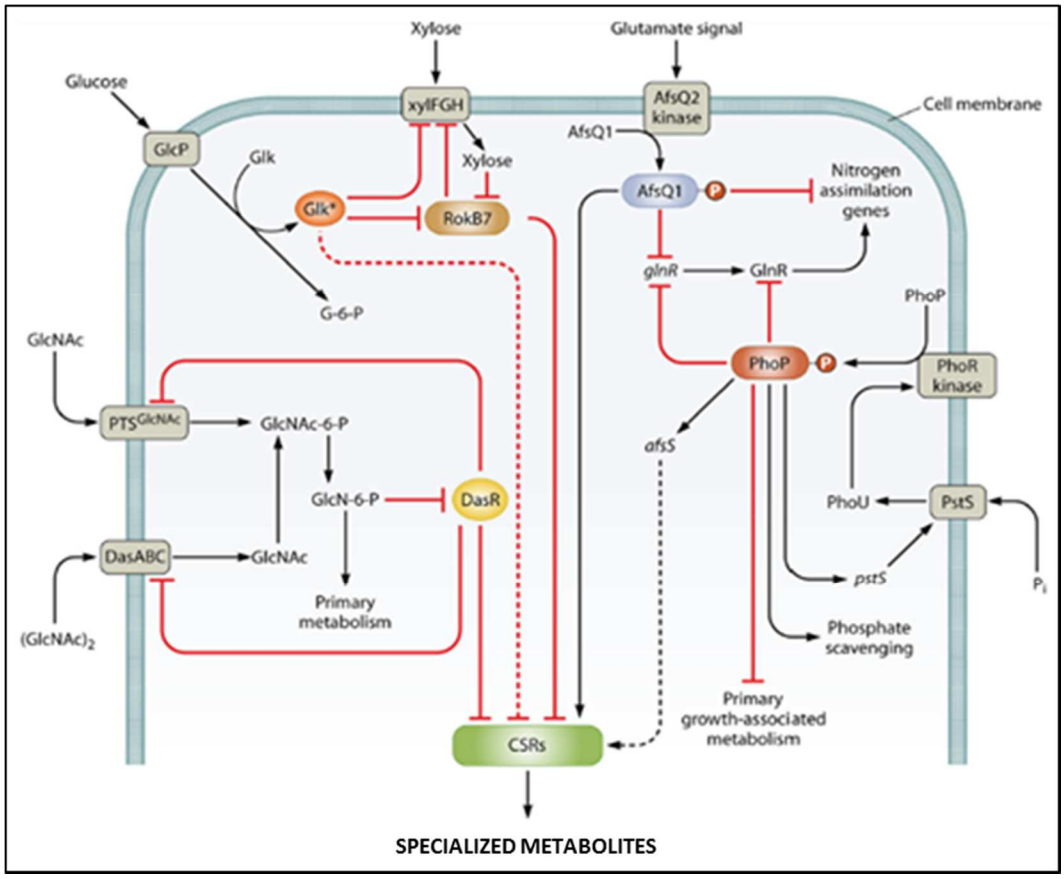

Figure 3 – The influence of the sources of carbon, nitrogen, and phosphate upon CSR expression. Solid lines represent direct interactions, while the dotted ones represent unknown routes (dotted lines). Arrows indicate activation, and bars indicate repression [Adapted from Liu *et al.*, 2013].

1.5 Aim of this work

The present work aimed to activate silent clusters in *Streptomyces natalensis* by manipulating CSR encoding genes.

Streptomyces natalensis ATCC 27448 was used as work model. This strain was first isolated from a soil sample originated from South Africa (Natal province) and is the producer of the antifungal polyene macrolide pimaricin (commercial name natamycin, E235) widely used in the food industry to prevent mould contamination of cheese and other non-sterile foods. Additionally, *S. natalensis* Δ*pimM*, a strain that lacks pimaricin production due to the absence of the positive CSR PimM (Anton *et al.*, 2007) has also been used.

As specific objectives it was proposed (1) the *in silico* and (2) transcription characterization of six biosynthetic gene clusters detected by antiSMASH in *S. natalensis* genome; (3) the overexpression of three selected CSR; and (4) evaluate the effect of CSR overexpression on the transcription and metabolic profile.

2. Materials & methods

2.1 Strains, growth conditions and plasmids

2.1.1 *Streptomyces* strains

The *Streptomyces* strains used in this work were *Streptomyces natalensis* ATCC 27448, a pimaricin producer, and the derivative *S. natalensis* 40D9-1 (Δ*pimM*) (Antón *et al.*, 2007), a pimaricin non-producing strain.

Sporulation was achieved in TBO medium [2% (w/v) oat flakes, 2% (w/v) tomato paste, 2.5% (w/v) agar, pH 6.5] (Higgens *et al.*, 1974) at 28 °C. After 8-10 days of growth, spores were scrapped from solid medium using a 0.025% (v/v) Triton X-100 and 30% (v/v) Glycerol solution, filtered with cotton and stored at -80 °C in the same solution. Spore concentration was determined spectrophotometrically at 600 nm using the following ratio: 0.1 units of absorption corresponds to 10^8 spores per mL (Kieser *et al.*, 2000).

For *Streptomyces* liquid cultures, 10^8 spores were inoculated in YEME medium without sucrose [0.3% (w/v) yeast extract, 0.5% (w/v) bacto peptone, 0.3% (w/v) malt extract, 5 mM $MgCl_2$ and 1% (w/v) glucose] (Kieser *et al.*, 2000) and grew in an orbital incubator shaker at 28 °C and 220 rpm, with a ratio of medium culture volume:flask volume of 1:10.

2.1.2 *Escherichia coli* strains

Escherichia coli ET12567 was used as a donor strain in intergeneric conjugation with *Streptomyces*. This strain does not methylate DNA (*dam*$^-$) (MacNeil, 1992), which helps circumventing the methyl-specific restriction system of *Streptomyces*. In addition, it harbours the non-transmissible plasmid pUZ8002 that contains the genes (*tra* genes) necessary for the mobilization of other vectors. Moreover, pUZ8002 also holds kanamycin and chloramphenicol as selection markers. *E. coli* DH5α was routinely used to transform and replicate plasmid DNA and also for bioassay experiments.

E. coli strains were routinely grown in LB medium [1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl] (Miller, 1972) at 37 °C. For solid LB medium 2% (w/v) agar was added.

2.1.3 Test microorganisms

The Gram-positive bacterial strains *Bacillus cereus* ATCC 14579, *Enterococcus faecalis* ATCC 29212 and *Staphylococcus aureaus* ATCC 29213 were used together with *E. coli* DH5α and *Saccharomyces cerevisiae* BY4741, as test organisms for bioassay experiments. These microorganisms were first grown O/N in submerged cultures and 1 mL was used for growth inhibition assays. *B. cereus* was grown in LB medium at 37 °C, *Enterococcus faecalis* and *Staphylococcus aureaus* were grown in TSB medium (Sambrook and Russel, 2001) at 37 °C and

Saccharomyces cerevisiae was grown in YPD medium [1% (w/v) yeast extract, 2% (w/v) bacto peptone, 2% (w/v) glucose] (Ausubel *et al.*, 1992) at 26 °C. For solid TSB and YPD media 2% (w/v) agar was added.

2.1.4 Plasmids

The *oriT*-containing pIB139 plasmid (Wilkinson *et al.*, 2002) was used for the expression of genes under the strong promoter *ermE**p in *Streptomyces* (Paget *et al.*, 1999). Apramycin stands as the selection marker for this plasmid.

PCR amplicons were cloned in pGEM®-T Easy vector (Promega) for DNA sequencing. Ampicillin is the selection marker for this plasmid, together with *lacZ* gene that codes for the β-galactosidase.

2.2 DNA procedures

2.2.1 Purification of total DNA from *Streptomyces*

Purification of total DNA from *Streptomyces* was carried out using the MasterPure™ Gram Positive DNA Purification Kit (Epicentre). A sample of 1 mL of culture broth was harvested and washed in 1 mL of TE buffer [10 mM Tris-HCl, pH 7.5, 1 mM EDTA]. For cell lysis, mycelium was resuspended in 300 µL of TE Buffer containing 2 µL of Ready-Lyse Lysozyme to each sample and incubated at 37 °C for 1.5 h. Then, 300 µL of Lysis Solution with 2 µL Proteinase K were added to each sample, mixed thoroughly and incubated at 65 °C for 30 min. DNA cleaning and purification was achieved firstly by adding 350 µL of MPC Protein Precipitation Reagent and mixing vigorously by vortex. Next, 2 µL RNAse A were added and samples were incubated at 37 °C for 30 min. Protein precipitation was carried out by mixing the samples with 1 vol. of phenol-chloroform-isoamyl alcohol (25:24:1). The aqueous phase was recovered after centrifugation (10 min, 12100 g) to a new tube and 0.6 vol. of isopropanol and 0.1 vol. of 3 M sodium acetate pH 5.2 were added. Samples were centrifuged at room temperature (10 min, 12100 g) and the supernatant was discarded. The DNA pellet was rinsed with 500 µL cold ethanol 75% (v/v) and resuspended in 50 µL of H_2O.

Plasmid DNA was isolated using the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, Saint Louis, MO) following manufacturers' instructions.

2.2.2 Enzymatic manipulation of DNA

For the ligation of DNA molecules a ligase isolated from the phage T4 (T4 DNA ligase) was used. T4 DNA ligase from Promega and Fermentas were used according to manufacturers' instructions.

Restriction endonucleases were used to cleave DNA of specific nucleotide sequences. *Eco*RV, *Nde*I, *Bgl*II, *Age*I, *Nco*I and *Not*I (Fermentas) were used according to the indications provided by the manufacturer.

2.2.3 Transformation of *E. coli* strains

E. coli DH5α transformation was performed as previously described (Sambrook and Russel, 2001). First, 5 µL of recombinant DNA was added to competent cells prepared by the rubidium chloride method and incubated on ice 45 min. Aliquots were then heat shocked at 42 °C for 1.5 min, 1 mL of LB was added, and cells were further incubated at 37 °C. Finally, recombinant cells were plated on solid LB medium, supplemented with IPTG (100 µM), X-Gal (20 µg/mL) and ampicillin (50 µg/mL) if transformed with pGEM®-T Easy derivative plasmids, or with apramycin (SIGMA-Aldrich) (50 µg/mL) if transformed with pIB139 derivative plasmids.

E. coli ET12567 [pUZ8002] was transformed by electroporation. Briefly, 1 mL of an O/N culture was inoculated in 10 mL of fresh LB supplemented with kanamycin (50 µg/mL), and chloramphenicol (25 µg/mL) and grown at 37 °C until an OD_{600nm} of 0.8-0.9. Then, cells were washed twice in 10 mL of cold-ice H_2O and in 5 mL of cold-ice 10% (w/v) glycerol. Cells were resuspended in 50 mL of 10% (w/v) glycerol and 100 ng of DNA was added. Electroporation was performed in a Gene Pulser® (BioRad) with an intensity of 2.5 Ω. Cells were left to recover at 37 °C for 1 h and plated in solid LB medium supplemented with kanamycin (50 µg/ml), chloramphenicol (25 µg/mL) and apramycin (50 µg/mL).

2.2.4 Intergeneric conjugation

A culture of the donor strain *E. coli* ET12567 [pUZ8002], harbouring pIB139 or derivatives was inoculated in LB supplemented with kanamycin (50 µg/mL), chloramphenicol (25 µg/mL) and apramycin (50 µg/mL), and grown overnight at 37 °C. 1 mL of this pre-culture was inoculated in 10 mL fresh LB supplemented with the same antibiotics and grown at 37 °C until an OD_{600nm} of 0.5-0.6. Cells were washed twice with LB medium without antibiotics and resuspended in 25 mL of the same medium. Spores of *S. natalensis* or *S. natalensis* Δ*pimM* were used as recipients. For each strain, 10^8 spores were washed with 2xTY medium [1.6% (w/v) bacto-tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl] (Sambrook and Russel, 2001) and resuspended in 500 µL of the same medium. Spores were then subjected to a heat-shock (45 °C for 10 min). 500 µL of *E. coli* ET1256 [pUZ78002] cells harbouring the plasmid to be transferred was added to the treated spores and the mixture was incubated at 30 °C for 15 min. The mixture was pelleted, resuspended in the residual liquid and spread on MS [2% (w/v) manitol, 2% (w/v) soya flour, 2% (w/v) agar %] (Hobbs *et al.*, 1989) supplemented with 10 mM $MgCl_2$. Plates were incubated at 30 °C for 16-18 h and then overlaid with 500 µL of an aqueous

solution containing apramycin (50 mg/mL) and nalidixic acid (25 mg/mL). Plates were incubated for further 10-15 days until exconjugants appeared. Exconjugants colonies were transferred to a new TBO plate supplemented with nalidixic acid and apramycin. For validation purposes, the identity of all new strains was confirmed by Southern blot hybridization and PCR.

2.2.5 Southern blot hybridization

The screening of the pIB139 plasmid integration into *S. natalensis* ATCC 27448 and *S. natalensis* 40D9-1 was made first by antibiotic resistance selection and confirmed by Southern blot hybridization, as previously described (Sambrook and Russel, 2001).

Initially, DNA was digested by restriction endonucleases and the resulting fragments were separated by electrophoresis in a 0.8% (w/v) agarose gel. The lambda (λ) phage DNA digested with *Hind* III labelled with digoxigenin (DIG) (Roche) was used as molecular weight ladder. DNA was transferred from the agarose gel to a nylon membrane (Amersham) in the vacuum system (Bio-Rad) and following manufactures' instructions.

DNA probes were labelled with digoxygenin using the "DIG-High prime" (Roche) kit and following the protocol provided by the manufacturer. Generally, 300-500 ng of DNA template were used in the digoxigenin labelling procedure. Hybridization and detection were performed as afore described (Sambrook and Russel, 2001) and using the alkaline phosphatase substrate CPD-Star (Roche).

2.3 Purification of total RNA from *Streptomyces*

Total RNA purification from *Streptomyces* was performed using RNeasy Mini Kit (Qiagen) and following a protocol optimized for *Streptomyces* (Beites *et al*., 2011). Briefly, 500 µL of *S. natelensis* culture broth was collected and mixed with 1 mL of RNA Protect Bacteria Reagent (Quiagen), incubated for 5 min at room temperature, harvested by centrifugation (10 min, 3600 g) and immediately stored at -80 °C. For total RNA extraction, mycelium pellets were resuspended in Buffer RLT with 1% (v/v) β-mercaptoethanol. Cellular disruption was achieved by sonication (Branson Sonifier, Model B-15) using cycles of 5 seconds, duty cycle of 50%and an output of 3. RNeasy Mini Spin columns were used to recover total RNA and DNA was removed by two serial DNase treatments, an in-column DNase I RNase-free (Qiagen), followed by a batch treatment using the DNA-free Kit (Ambion). The absence of genomic DNA was confirmed by PCR.

RNA quantification was assessed by NanoDrop ND-1000 spectrophotometer (Thermo Scientific), and its quality and integrity was evaluated in an Experion™ Automated Electrophoresis System (Bio-Rad).

2.4 Polymerase Chain Reactions (PCR) and Reverse Transcription (RT) PCR

DNA amplification by PCR was performed with Pfu DNA polymerase (Fermentas) or GoTaq Flexi DNA Polymerase (Promega) according to manufactures' instructions. The DNA amplification was performed in a C1000™ Thermal Cycler (BIORAD). Used primers are listed in Table 1.

The genetic identity of the PCR products used for overexpression was confirmed by sequencing (STAB Vida). For gene expression studies 1 µg of DNase-treated total RNA was transcribed with the iScript™ Select cDNA Synthesis Kit (Bio-Rad), following the manufacturer's instructions. To check if there was any trace of genomic DNA in the samples, a standard PCR reaction was performed using specific primers for *S. natalensis* ATCC 27448 16s rRNA gene and 3 µl of the DNAse treated RNA sample as template. As a positive control it was used the *S. natalensis* ATCC 27448 genomic DNA. The housekeeping gene 16S rRNA was used as a positive internal control for both strains, while the PimM direct target *pimS1* (pimaricin biosynthetic gene cluster) (Santos-Aberturas *et al.*, 2011) was used as a negative control for the Δ*pimM* strains. Genomic DNA was used as control for RT-PCR reactions where transcripts were not detected, or detected in low levels.

Table 1 – Primers used in this work. They are categorized according to their application: RT-PCR, gene overexpression or strain validation. Primers included in the strain validation category were used for both PCR studies and amplification of the Southern blot hybridization probes. Primers marked with (*) were also utilized for strains validation by PCR. Primers marked with (†) were additionally used for the development of the probes required for Southern blot hybridization.

Primer	Sequence (5'-3')	Target	Amplicon Size (bp)
RT-PCR			
S1C1R1_AS	ACCAACGACCAGGACACCCTC	*SNA00088*	119
S1C1R1_S	TCGCGCAGGGAACGGAAG		
S1C1R2_AS	ATGCGAACCGGCGAGGAC	*SNA00096*	102
S1C1R2_S	GCTGACCCCCATGAAGCACA		
S1C1S1_AS	TCAAGAGCCCCGGCCGAC	*SNA00095*	104
S1C1S1_S	GTCGTTGTGGACCTCTACCGCC		
S1C2R1_AS	AAGTGGACCTGGACGGCGAG	*SNA00399*	118
S1C2R1_S	TGGAAGACCATCCGGACCTCT		
S1C2R2_AS	GAGTGCGTCGAAGCCATGCG	*SNA00403*	125
S1C2R2_S	GCCTTCGCGCCTTCGATGAA		
S1C2R3_AS	GGGCCGTGATGGACGAGACC	*SNA00404*	133
S1C2R3_S	CCGGAAGCGAAGGGCATGAC		
S1C2R4_AS	TCATGGACGGTCTGCTGCTG	*SNA00428*	152
S1C2R4_S	CTTGGGACGGTAGGGGTGGT		
S1C2R5_AS	ACCTGGTCGTCTTTCTGCGC	*SNA00433*	115
S1C2R5_S	TTCTCGGTGATCCACTGCTGG		

Table 1 – Primers used in this work (continuation).

Primer	Sequence (5'-3')	Target	Amplicon Size (bp)
RT-PCR			
S1C2S1_AS	GACACACCGCGACCTTCACC	*SNA00397*	122
S1C2S1_S	GAAGACCACATCGACCTCCTCC		
S1C2S2_AS	ACGACCGTACCGCCGACAT	*SNA00414*	157
S1C2S2_S	ACGACCGGCAGCGACTTCT		
S1C3R1.2_AS	AGACGCACGGCCTGCTACG	*SNA00830*	117
S1C3R1.2_S	TGTAGGTCTGGACGACGATGCC		
S1C3R2.2_AS	TCGTAGCCGGTCAGCAGCC	*SNA00835*	141
S1C3R2.2_S	ACACCTCGACGTCCCCGTAC		
S1C3R3_AS	ATTCCCAAGGGGTCGGTGA	*SNA00851*	193
S1C3R3_S	GCACCACGAAAGGCAGCTT		
S1C3S1_AS	CCGGCGACCTGGACTTCTACAT	*SNA00841*	133
S1C3S1_S	GCGATGTTGCCGTACTCGGTC		
S1C4R1_AS*; †	AGAGTGCGAGGCCGTTGGG	*SNA01173*	106
S1C4R1_S †	AACCCGCGCGACATCTGC		
S1C4S1_AS	ATGGCGCTGGGTGTCGGT	*SNA01166*	115
S1C4S1_S	CGCCCGCGCCTATGACTT		
S1C5R1_AS	CAGGAACTCCCGGAAGATGCC	*SNA01232*	174
S1C5R1_S	AAGGTCGGCAAGGGAACGGT		
S1C5R2_AS*; †	CGCAGAACACTCCGACGCA	*SNA01239*	102
S1C5R2_S †	GGTCATGGACCGGGAGCTCT		
S1C5R3_AS	ATCTCGCGGCAGTCCGGTC	*SNA01250*	103
S1C5R3_S	CGTGGTGGCCCAGGAAGATG		
S1C5R4_AS	ATCGACGTACGAGGCGAGCT	*SNA01254*	147
S1C5R4_S	GAGCGGCACAAGCGTCATCT		
S1C5R5_AS	GATCCTCGCTGAAGGTCGTCTC	*SNA01258*	109
S1C5R5_S	CTCCAGGACGCCATCGACTT		
S1C5R6.2_AS	GGACTGGACCCGCTCGAAGA	*SNA01256*	143
S1C5R6.2_S	CCAAGGACGACCTCACCCTC		
S1C5S1_AS	GGCAGCAGGTCTACGAGCGTAT	*SNA01225*	117
S1C5S1_S	ACTGCCGCATCTCTGCCAGA		
S5C20R1_AS	TGACCGTCCAGGGCACTGAAC	*SNA06239*	108
S5C20R1_S	CTTCCGCAGCCACCCATGG		
S5C20R2_AS	ACGGCGAGGCCCTGAATGTAG	*SNA06240*	132
S5C20R2_S	CCCCTTGATGCTGCCCAA		
S5C20R3_AS	TGAGCTGGAGGCGACTGTGCT	*SNA06245*	157
S5C20R3_S	AAGGATTTCGCCGCCGGA		
S5C20R4_AS*; †	CTTTCCCGGCAGCGATGA	*SNA06246*	157
S5C20R4_S †	CCATCTGGGAAATGACGCGC		
S5C20R5_AS	ACGATGCCCTTCCCGCTGA	*SNA06261*	114
S5C20R5_S	CTCCGGTTTCTGTTTCTCGCG		

Table 1 – Primers used in this work (continuation).

Primer	Sequence (5'-3')	Target	Amplicon Size (bp)
RT-PCR			
S5C20R6_AS	CGGCAGCCGATGGTGAGTT	*SNA06243*	192
S5C20R6_S	CCTCATCCGCGCCTACACG		
S5C20S1_AS	TACTCACCATGGCTGTGCCCC	*SNA06257*	160
S5C20S1_S	TGCTGTCCGTGCTGTTCCCT		
PimS1_AS	CCACCTTCGGACGCAACAC	*pimS1*	108
PimS1_S	TGGCTCGGCATCGGATCA		
16S_AS	CTCCTCAGCGTCAGTATCG	16s rRNA	117
16S_S	CAGGCTAGAGTTCGGTAGG		
Gene overexpression			
SNA01239_F	TAGGGCATATGGACGCGATGACCCTCTAC	*SNA01239*	957
SNA01239_R	AGGATATCGACTATACGTCGCGGTCGATCCG		
SNA06246_F	GCGGCATATGAAGAACTCACTGCTGGTC	*SNA06246*	672
SNA06246_R	CCGATATCTCAGCCGGGGAGGAGGATCTG		
SNA01173_F	GGCCCATATGGTCGTCGTGGCCGAC	*SNA01173*	617
SNA01173_R	GCGATATCGGTCAGTGGATGATGCCCGTACG		
Strains validation			
pIB_U_F	GGCTTGCGCCCGATGCTAGT	pIB139 plasmid	293
pIB_U_R	CCGGCTCGTATGTTGTGTGG		

2.5 Bioassay experiments

S. natalensis strains were cultured in YEME liquid medium for 72 h, in order to ensure the production of secondary metabolites. 8 mL of culture broth were centrifuged twice (5 min, 12 100 g) and 250 µL of the recovered supernatants were concentrated, until a final volume of 40 µL, before being transferred to holes of 0.5-cm diameter cut in solid LB, TSB or YPD plates, and let diffuse in the media for 1 h at room temperature. Next, 1 mL of O/N cultures of test microorganisms was added to the plates, within 15 mL of pre-warmed LB, TDB or YPD supplemented with 1% (w/v) agar. Inhibition zones were measured 2 days later using a Vernier caliper.

2.6 Analysis of specialized metabolites by high performance liquid chromatography (HPLC)

For pimaricin quantification, 500 µL of *S. natalensis* culture broth was collected and mixed with 1 vol. of methanol. Samples were left 2 h at 30 °C with agitation in order to enhance the extraction of the pimaricin to the organic phase. The organic phase was recovered through centrifugation (5 min, 12 100 g) before HPLC analysis. Pimaricin quantification was performed on a Merck-Hitachi liquid chromatograph (Merck-Hitachi, Tokyo, Japan), equipped

with a detector L-4000, an autosampler L-7250, a pump L-7100, an L-7300 column oven and an interface D-7000. Chromatograms were recorded and processed on a Merck-Hitachi D-7000 HSM PC-based chromatography data system, setting wavelength at 304 nm. Strain metabolic profile was performed on a Dionex UltiMate 3000 system (Thermo Fisher Scientific, Bremen, Germany) equipped with an UltiMate 3000 RS pump, an UltiMate 3000 RS autosampler and an UltiMate 3000 RS column compartment, and coupled to an Accela PDA detector. Both systems were filted with a SunFire™ C_{18} column (5 µm, 4.6 × 250 mm; Waters). Elution was performed with a gradient of mobile phase composed of 0.1% (v/v) trifluoroacetic acid (solvent A) and HPLC-grade methanol (solvent B) at a flow rate of 0.6 mL/min and according to the following progam (methanol concentration): 50% B, 0-3 min; 90% B, 3-12 min; 90% B, 12-22 min; down to 50% B, 22-26 min; 50% B, 26-30 min. Purified pimaricin (Biocam™) was used as standard solution for the quantification of this metabolite.

Dry weight was used to determine the specific pimaricin production. 1 mL of culture broth was collected and washed twice in 1mL of 0.9% (w/v) NaCl. Mycelium was pelleted by centrifugation (5 min, 12 100 g) and dried at 60 °C for 5 days.

2.7 *In silico* analysis

2.7.1 Prediction and analysis of secondary metabolite biosynthetic gene clusters

Secondary metabolite gene clusters were identified *in silico* using the bioinformatic tool antiSMASH (http://antismash.secondarymetabolites.org/). Additionally, this software was also utilized for the prediction of the putative chemical structures of the products derived from PKS, NRPS or hybrid PKS-NRPS gene clusters, as well as the domain specificities of their biosynthetic core genes. Finally, in order to get more reliable predictions on the clusters' genes function and the chemical structure of the secondary metabolite, antiSMASH results have been manually curated. In specific, a BLASTp analysis (www.ncbi.nlm.nih.gov/BLAST/) on every gene of each cluster has been performed, as well as a literature review regarding putative identical biosynthetic gene clusters.

2.7.2 Internet resources

EMBL (European Molecular Biology Laboratory, Germany): www.ebi.ac.uk/.

BLAST: www.ncbi.nlm.nih.gov/BLAST.

NCBI genome database: www.ncbi.nlm.nih.gov/genome/browse.

Streptomyces genome project: http://streptomyces.org.uk/.

S. avermitilis genome project: http://avermitilis.ls.kitasato-u.ac.jp/.

3. Results and discussion

3.1 *S. natalensis* ATCC 27448 genome mining

The *S. natalensis* ATCC 27448 genome has been recently sequenced (Mendes *et al.*, unpublished results). The draft genome is distributed among 40 scaffolds and is estimated to have a total size of 8 653 788 bp, a G+C content of 70.63% and an estimated total of 8 318 protein-coding genes. The *S. natalensis* genome was submitted to a preliminary genome mining analysis using antiSMASH (http://antismash.secondarymetabolites.org/) – a software especially developed to identify and delimit secondary metabolites biosynthetic gene cluster (for more detailed description see section 1.3).

The antiSMASH analysis retrieved a total of 29 predicted biosynthetic gene clusters putatively involved in the production of secondary metabolites (a.k.a. specialized metabolites), including 8 terpenes, 3 bacteriocin-like toxins, 3 siderophores, 3 polyketides, 2 nonribosomal peptides, 1 oligossacharide, 1 phosphonate, 1 polyketide-siderophore-terpene hybrid and 1 polyketide-terpene hybrid (Table 2). Based on PFAM domain probabilities 33 additional putative biosynthetic gene clusters were identified with no assigned category. Moreover, the majority of the identified secondary metabolite biosynthetic gene clusters harboured at least one putative cluster situated regulator (CSR). Finally it should be mentioned that the 29 biosynthetic gene clusters may turn out to be 34 due to the classified hybrid ones (S1C2, S3C9 and S3C14). antiSMASH defines gene clusters by locating clusters of signature gene pHMM hits, spaced within <10 kb mutual distance, including flanking accessory genes. Then, gene clusters are extended by 5, 10 or 20 kb on each side of the last signature gene pHMM hit (Medema *et al.*, 2011). Consequently, gene clusters that are closely located may be merged and wrongly assigned as a hybrid one.

Due to time constraints, this work has focused on six clusters (S1C1, S1C2, S1C3, S1C4, S1C5 and S5C20). The selection process was based on (1) the chemical family of the predicted metabolite, to favour chemical diversity, (2) low genetic synteny with homologous gene clusters, to increase the probabilities of finding a new compound and (3) the existence of positive CSRs. Biosynthetic gene clusters were named according to the scaffold they belong sequentially (e.g. S3C12 is the 12th found cluster and is located in scaffold no 3). A detailed *in silico* analysis of the selected clusters was performed based on the antiSMASH results. Individual BLASTp results for each gene of the selected clusters are presented in Tables A1-6 (Appendix I). Table A7 summarizes the CSR found and their predicted role (i.e. if activators or repressors) based on previous studies. It should still be noted that all the proposed functions were foreseen according to (1) the position of the homologous hits presented by NCBI, (2) the

number of similar obtained homologies and (3) the core domains detected by the alignment search tool.

Table 2 – *S. natalensis* ATCC 27448 biosynthetic gene clusters predicted by antiSMASH. Gene *loci* was identified with the SNA prefix in accordance with the genome sequencing data.

Gene cluster	Type	No Scaffold: Location (bp)	SNA *loci*
S1C1	Terpene	S1: 77463-100227	*SNA00083-SNA00105*
S1C2	PKS-II-terpene	S1: 370687-434883	*SNA00375-SNA00437*
S1C3	PKS-III	S1: 835946-878512	*SNA00822-SNA00859*
S1C4	Bacteriocin	S1: 1186177-1211143	*SNA01156-SNA01179*
S1C5	NRPS	S1: 1222359- 1330255	*SNA01188-SNA01259*
S1C6	Siderophore	S1: 1490934-1508777	*SNA01393-SNA01404*
S2C7	Phosphonate	S2: 524293-566518	*SNA03120-SNA03161*
S3C8	Other	S3: 1795-46346	*SNA03856-SNA03898*
S3C9	Terpene-Lantipeptide-NRPS	S3: 278972-346052	*SNA04152-SNA04201*
S3C10	Terpene	S3: 390167-411216	*SNA04263-SNA04289*
S3C11	Terpene	S3: 492921-516009	*SNA04392-SNA04413*
S3C12	Oligosaccharide	S3: 519254-572380	*SNA04417-SNA04462*
S3C13	PKS-I	S3: 587907-651587	*SNA04479-SNA04537*
S3C14	PKS-I-Siderophore-Terpene	S3: 652887-799820	*SNA04539-SNA04613*
S3C15	Lantipeptide	S3 985892-1012524	*SNA04789-SNA04809*
S3C16	Bacteriocin	S3: 1126391-1138879	*SNA04917-SNA04925*
S4C17	Terpene	S4: 336772-360465	*SNA05344-SNA05364*
S4C18	Bacteriocin	S4: 574834-586218	*SNA05566-SNA05578*
S5C19	Terpene	S5: 66559-91614	*SNA05856-SNA05877*
S5C20	PKS-I	S5: 460920-515784	*SNA06231-SNA06280*
S6C21	NRPS	S6: 266965-367965	*SNA06673-SNA06733*
S6C22	Siderophore	S6: 491157-507468	*SNA06857-SNA06869*
S7C23	Other	S7: 100-42269	*SNA06906-SNA06955*
S7C24	Terpene	S7: 435422-456849	*SNA07423-SNA07440*
S8C25	Ectoine	S8: 393588-404835	*SNA07875-SNA07887*
S10C26	Terpene	S10: 267-15488	*SNA01458-SNA01477*
S10C27	Siderophore	S10: 301253-310365	*SNA01750-SNA01755*
S13C28	Terpene	S13: 20834-44430	*SNA02207-SNA02226*
S16C29	Butyrolactone	S16: 433-13110	*SNA02426-SNA02441*

Finally, the transcription of one selected structural gene together with CSRs encoding genes from the six clusters was analysed by RT-PCR nature. The large amount of structural genes, together with the short period of time available for this project, justifies the performed selection. For gene expression studies RNA samples were collected at early experimental phase where secondary metabolism is active. Two biological replicates were analysed.

3.1.1 S1C1 – a putative terpene biosynthetic gene cluster

3.1.1.1 *In silico* analysis

Cluster S1C1 spanning from 682408-700166 in scaffold 1 (*SNA00083-SNA00105*) was classified by antiSMASH as responsible for the production of a terpene (Fig. 4). Terpenes (a.k.a.

isoprenoids) are a ubiquitous group of molecules biosynthetically derived from units of isoprene precursor units (C_5H_8). Currently, over 55 000 of terpenes have been isolated and are present in our daily life, such as components of fragrances, hormones and medicines, among other products (Nicolaou & Montagnon, 2008; Chen & Baran, 2009).

In *Streptomyces*, terpenes biosynthesis can occur via two independent non-homologous metabolic pathways: the methylerythritol phosphate (MEP) pathway and the mevalonate (MVA) pathway (Dairi, 2005). Both pathways converge in the formation of isopentenyl pyrophosphate (IPP), from the condensation of a pyruvate and a glyceraldehyde-3-phosphate (G3P) (MEP pathway) or from the condensation of acetyl-CoA and acetoacetyl-CoA (MVA pathway) (Fig. 5).

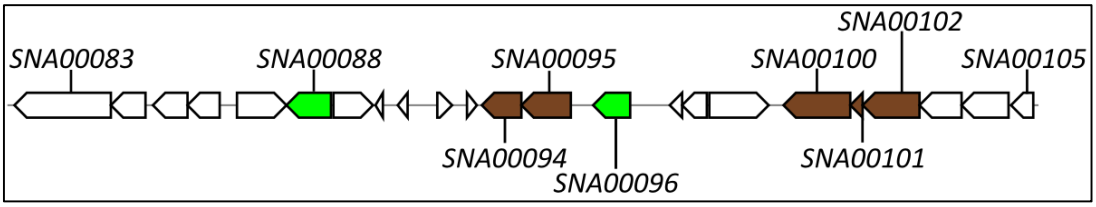

Figure 4 – Schematic representation of cluster S1C1. Brown arrows represent putative structural genes, which might be involved in the biosynthesis of the end terpene (SNA00094, methyltransferase; SNA00095, terpene synthase metal-binding domain-containing protein; SNA00100, AMP-dependent synthase SNA00101, phosphopantetheine-binding protein; SNA00102, decarboxylase). Green arrows represent the putative positive CSRs encoding genes: *SNA00088* (LysR-family); *SNA00096* (Cyclic AMP (cAMP) receptor protein). White arrows represent other genes.

Once formed, either by MVA or MEP pathways, IPP is rearranged by isopentenyl diphosphate:dimethylallyl diphosphate isomerase (IPP isomerase), to form dimethylallyl diphosphate (DMAPP) (Dairi, 2005).

The condensation of DMAPP with one, two or three IPP units gives rise to geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) that once cyclized by terpene synthases, constitute the parent skeletons of monoterpenes (C10), sesquiterpenes (C15) and diterpenes (C20). Two farnesyl units joined head-to-head by squalene synthase (SQS) lead to the formation of squalene, the precursor of triterpenes (C30). Tetraterpenes (C40) are formed from two geranylgeranyl units joined head-to-head by phytoene synthase (PSY), originating phytoene as an intermediate. Terpenes structural backbones can be further modified by hydroxylation, methylation and glycosylation, among others. It should be noted that terpene synthases are also referred as terpene cyclases (Greenhagen and Chappell, 2001).

BLASTp and protein domain analysis of SNA00095 identified a terpene synthase metal-binding domain-containing protein between amino acid 60 and 367 (PFAM03936) with 82% identity (306 aa/373 aa) with the one from *Streptomyces violaceusniger* Tu 4113 (Table A1) which corroborates the antiSMASH prediction. Furthermore, SNA00095 also shares 79%

identity (296 aa/374 aa) with a 2-methylisoborneol (MIB) synthase from *Streptomyces* sp. NRRL S-237. This enzyme is involved in the formation of the monoterpene MIB (Fig. 6A) – an organic chemical with a strong odour, which is one of the chemicals with major influence on the quality of drinking water (Newcombe and Cook, 2002). MIB biosynthesis is associated to clusters that include, besides the monoterpene cyclase, a cyclic nucleotide-binding protein and a methyltransferase, which are both also present in S1C1: SNA00096 [73% identity (52 aa/71 aa) with *Streptomyces* sp. NRRL S-237] and SNA00094 [88% identity (261 aa/ 265 aa) with *Streptomyces* sp. NRRL S-237], respectively (Fig. 6B). The results suggested that S1C1 may be responsible for the production of a MIB analogue or structurally related compound.

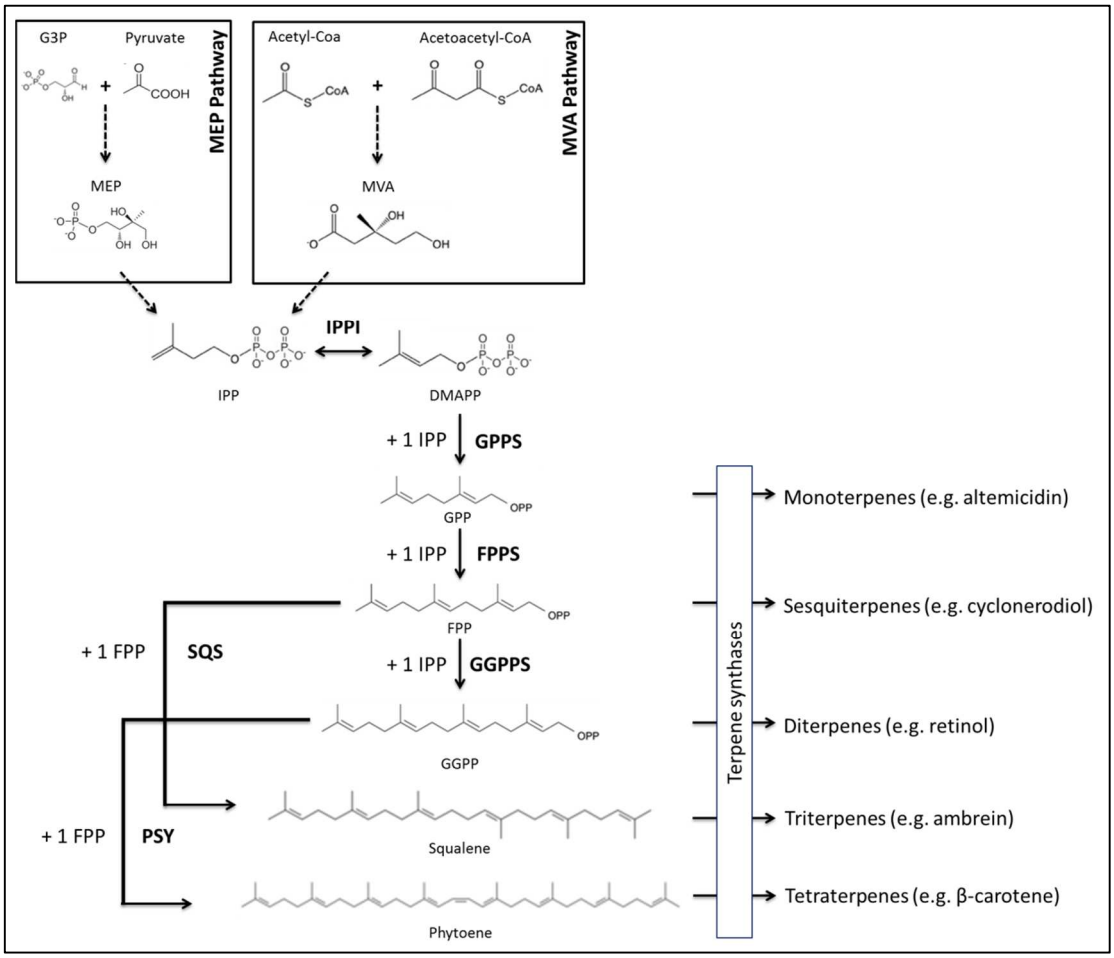

Figure 5 – Schematic representation of terpenes biosynthesis. Once formed either from MEP or MVA pathways, IPP is converted to its allylic isomer DMAPP, by IPPI. The condensation of one, two or three molecules of IPP to DMAPP, by GPPS, FPPS or GGPP, respectively, lead to the formation of GPP, FPP or GGPP. In turn, two molecules of FPP, by SQS, or GGPP, by PSY give rise up to squalene or phytoene respectively. The latter compounds, as well as GPP, FPP and GGPP, once cyclised by terpene synthases give rise to monoterpenes, sesquiterpenes or diterpenes, respectively. (G3P, glyceraldehyde-3-phosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate; MVA, mevalonate; IPP, isopentenyl diphosphate; IPPI, isopentenyl diphosphate isomerase; DMAPP, dimethylallyl diphosphate; GPPS, geranyl diphosphate synthase; GPP, geranyl diphosphate; FPPS, farnesyl diphosphate synthase; FPP, farnesyl diphosphate; GGPPS, geranylgeranyl diphosphate synthase; GGPP, geranylgeranyl diphosphate; SQS, squalene synthase; PSY, phytoene synthase) [Adapted from Schmidt *et al.*, 2010].

Flanking the core S1C1 region (*SNA00094-SNA00096*) it can be identified putative tailoring enzymes encoding genes, i.e., enzymes that further functionalize the terpene molecule, particularly SNA00100, SNA00101 and SNA00102. SNA00102 was identified as a putative decarboxylase. These enzymes, also known as carboxy-lyases, are able to decarbolxylate their substrates, releasing carbon dioxide (Ochoa, 1951). They have already been shown to be responsible for generating structural diversity in some cases during the maturation of secondary metabolites (Rachid *et al.*, 2010).

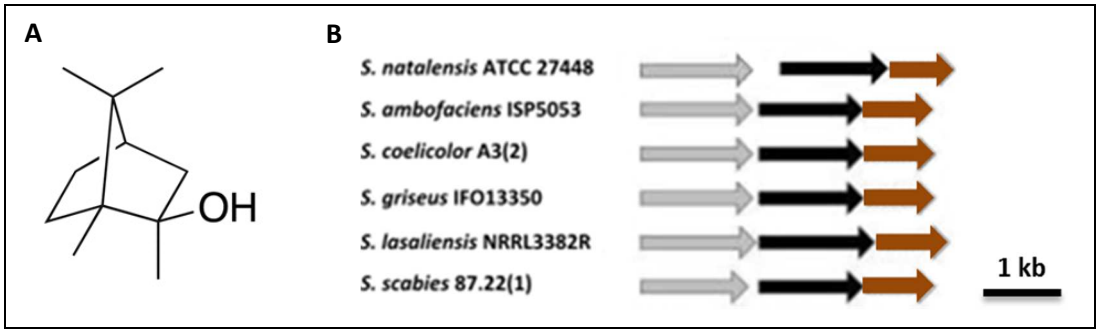

Figure 6 – 2-methylisoborneol (MIB). A). Chemical structure. B). Organization of the biosynthetic gene cluster, including the predicted monoterpene cyclase (black arrows) and flanking genes in different *Streptomyces* strains. Greyed and brown arrows represent cyclic nucleotide-binding protein and methyltransferase genes, respectively [Adapted from Komatsu *et al.*, 2008].

Regarding CSR proteins, S1C1 presented two putative positive transcriptional regulators: SNA00088 from the LysR family and SNA00096, a cAMP receptor protein (CRP)/fumarate and nitrate reduction regulatory protein (FNR)-type transcription factor (Table A7). SNA0096 was not identified as a regulator by antiSMASH. However the *in silico* analysis performed has shown that it is part of the terpene biosynthetic cluster.

3.1.1.2 Assessing the transcription of selected genes of S1C1 biosynthetic gene cluster

Cluster S1C1 holds two putative positive transcription regulators: SNA00088, from the LysR-family, and SNA00096, from CRP/FNR-family. The transcription of both encoding genes was assessed together with *SNA00095*, a structural gene that encodes a putative terpene synthase (Fig. 7). RT-PCR results showed that the putative transcription regulator encoding gene *SNA00096* is being transcribed in both strains, as well as the putative structural gene *SNA00095*. Conversely, *SNA00088* is being poorly expressed in both wild-type and Δ*pimM* strains. Regarding the differences between strains, in spite of the variations observed in biological replicates, it seems that the transcript levels of *SNA00095* and *SNA00096* are higher in Δ*pimM*.

These results pointed out S1C1 as a non-silenced gene cluster because the structural gene *SNA00095* is being expressed. Nevertheless production of terpenes has not been described in *S. natalensis*. To overproduce this metabolite, overexpressing *SNA00096* might be

strategy, since it was predicted by the *in silico* analysis as a putative positive regulator and its expression pattern is similar to the one presented by *SNA00095*. The low transcription levels of *SNA00088*, another predicted positive CSR, suggest that it is in a different regulatory lawyer. Therefore, its overexpression could also work as a viable approach for increasing the terpene production.

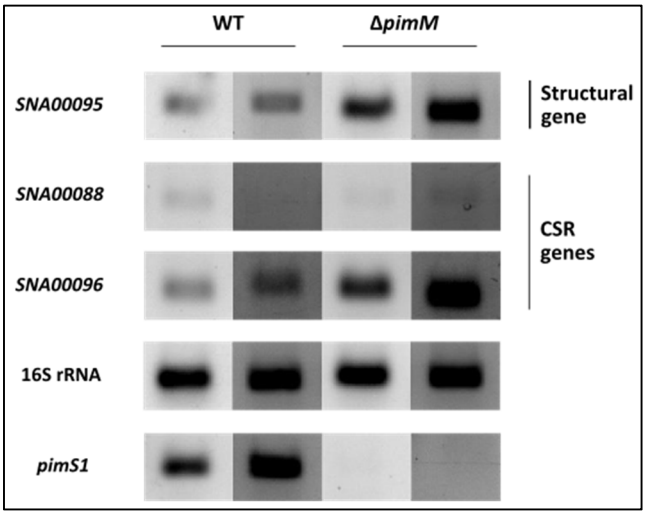

Figure 7 – Transcription analysis of S1C1 selected genes by RT-PCR. Transcripts of *SNA00095* (terpene synthase encoding gene), *SNA00088* (LysR-family encoding gene) and *SNA00096* (CRP/FNR-family encoding gene), 16S rRNA (positive control) and *pimS1* (positive control for WT and negative control for Δ*pimM*) were analysed by RT-PCR with 30 cycles on cDNA generated from RNA isolated at early stationary phase. Two biological replicates are shown for each strain.

3.1.2 S1C2 – a putative type II PK-terpene hybrid biosynthetic gene cluster

3.1.2.1 *In silico* analysis

Cluster S1C2 was identified by antiSMASH as a hybrid type II PKS-terpene (Fig. 9), meaning that the final compound may have mixed characteristics between a type II polyketide and a terpene. Polyketides are a large family of structurally diverse natural products that possess a wide range of pharmacological properties (e.g. antibacterial, antitumor, insecticide, among others) and display a vital role in human and veterinary medicine (Moore and Hertweck, 2002). The production of these compounds share common mechanisms with fatty acids biosynthesis. Indeed, each repetitive elongation step involves a Claisen condensation, where acyl extender units are incorporated into the growing chain. However, in fatty acids biosynthesis each elongation step is followed by a complete cycle of reductive reactions, such as ketoreduction, dehydration, and enoylreduction which may or not occur in polyketides synthesis, leading to the formation of far more chemically diverse products (Aparicio *et al.*, 2002). The backbone of polyketides is synthesised by polyketide synthases, which can be categorized into: type I (PKS-I), type II (PKS-II) – majorly found in microorganisms – and type III PKS (PKS-III) – more common in plants, even though some *Streptomyces* strains harbour genes encoding for this kind of enzymes (Yu *et al.*, 2012). PKS-I, also known as modular PKSs, are

multifunctional and multimodular proteins that are mostly involved in the synthesis of macrolides. Each PKS-I is composed by at least one module with three core catalytic domains (minimal module): acyltransferase (AT), acyl carrier protein (ACP) and ketosynthase (KS). This minimal module is responsible for an elongation cycle, i.e. incorporation of an extender unit (e.g. malonyl-CoA). In addition to these catalytic domains, PKS-I modules may also contain domains that alter the reduction state of the growing chain in a NADPH-dependent manner, such as β-ketoreductase (KR), dehydratase (DH) and enoylreductase (ER). When the polyketide growing chain reaches the final module, elongation is terminated and the molecule is released by hydrolysis through the action of a thioesterase domain (TE). In contrast to the type I PKS, type II PKS, or iterative PKS, are multienzymatic complexes mainly involved in the production of aromatic polyketides (e.g. tetracyclines and actinorhodin). These enzymes elongate polyketides through repetitive condensation reactions catalysed by heterodimers composed by ketosynthase and ACP, in an iterative way. Furthermore, PKS-II presents a high degree of substrate specificity, mostly using acetyl-CoA and malonyl-CoA as starter and extender units, respectively. After the condensation procedure, the aromatic polyketides can be reduced by ketoreductase and modified by enzymes like aromatases, cyclases or glycosylases (Hwang *et al.*, 2014).

Regarding cluster synteny, S1C2 ClusterBlast analysis revealed two regions (*SNA00394-SNA00402* and *SNA00409-SNA00435*) that shared high similarity and synteny with homologous gene clusters from *Streptomyces* (Fig. 8), i.e. homologous genes have identical co-localization within the different genomes of the different species. Individual BLASTp analysis for each gene unveiled that the conserved regions hold core genes responsible for the production of a terpene (*SNA00409-SNA00435*) and for a type II PKS (*SNA00394-SNA00402*) (Fig. 9; Table A2). It should be noted that core genes are intended to be specifically related with the biosynthesis of the backbone of the corresponding metabolite.

BLASTp analysis of each gene within the type II PKS conserved region (*SNA00393-SNA00905*) revealed the presence of the three enzymes that constitute a PKS minimal module: SNA00397, a putative polyketide chain length factor, SNA00398, a putative polyketide beta-ketoacyl synthase and SNA00396, a putative ACP (Table A2). Genes coding for these proteins and surrounding ones (*SNA00393-SNA00401*) showed high degree of identity with the *whi* locus of *S. coelicolor* (Fig. 10) involved in the production of the polyketide spore-associated grey pigment. Protein domain analysis identified a putative SchA/CurD domain (PFAM04486) in SNA00400, located between aa 137-251. These domain containing proteins function is unknown, but they have been identified as be part of gene clusters involved in the synthesis of polyketide-based spore pigments (Blanco *et al.*, 1993).

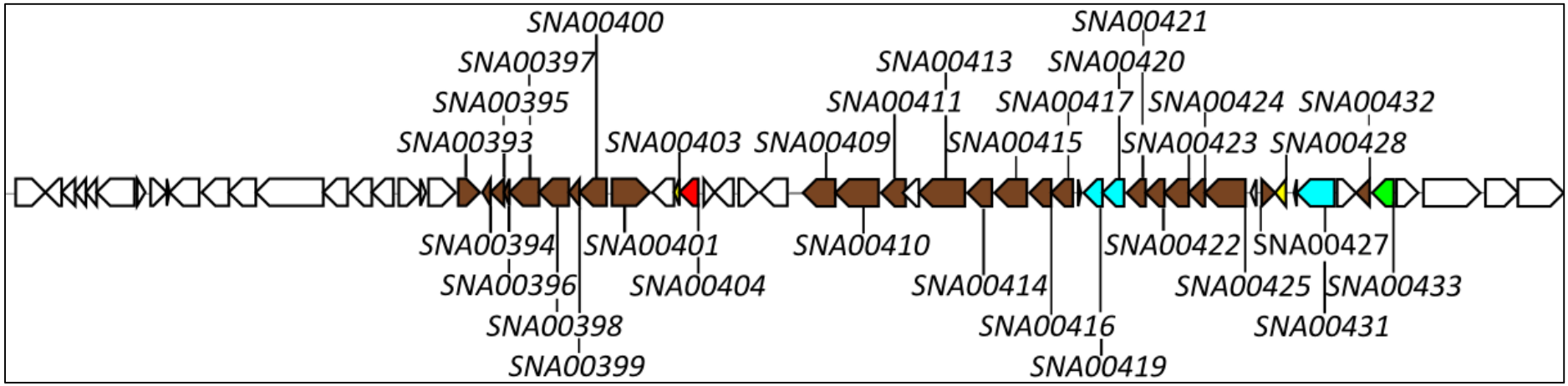

Figure 8 – Schematic representation of cluster S1C2. Brown arrows represent putative structural genes. The upper ones (from *SNA00393* to *SNA00401*) are likely to be involved in the biosynthesis of a polyketide, while the other lower ones are likely to be involved in the biosynthetic pathway of a terpene (from *SNA00409* to *SNA00432*). Blue arrows represent putative transporters encoding genes: *SNA00419*, *SNA00420* and *SNA00431*. The green arrow represents the putative positive CSR encoding genes: *SNA00432* (AraC-family). The red arrow represents a putative negative transcription regulator encoding gene: *SNA00404* (XRE-family). Yellow arrows represent genes that code for putative CSRs which may either act as activators or repressors: *SNA00403* (AbaA-like protein); and *SNA00428* (putative regulatory protein). White arrows represent other genes.

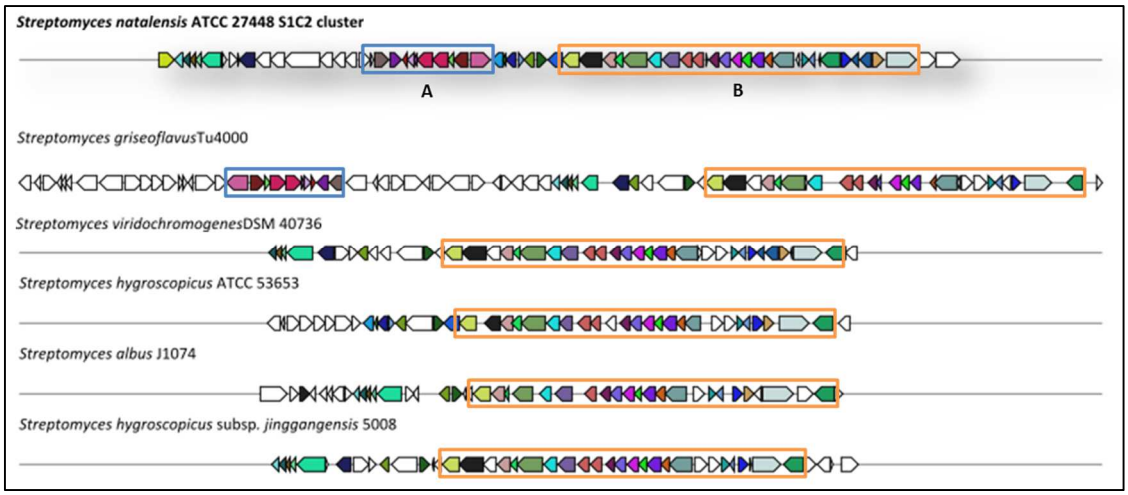

Figure 9 – Top five S1C2 homologous gene clusters regarding synteny. Homologous genes are shown in identical colours, whereas white-coloured genes have no Blast hits between the gene clusters. Putative PKS-II biosynthetic core genes are marked within blue boxes (A), while putative terpene biosynthetic genes are marked within orange boxes (B). In *S. natalensis* ATCC 27448 the dashed box is localized between *SNA00392* and *SNA00401*, while the non-dashed is localized between *SNA00409* and *SNA00435*. Accession numbers: *S. griseoflavus* Tu4000 – NZ_GG657758.1_c4; *S. viridochromogenes* DSM 40736; *S. hygroscopicus* ATCC 53653 – NZ_GG657757.1_c24; *S. albus* J1074 – NZ_GG657754.1_c12; *S. hygroscopicus* subsp. *Jingangensis* 5008 – NZ_DS999645.1_c18.

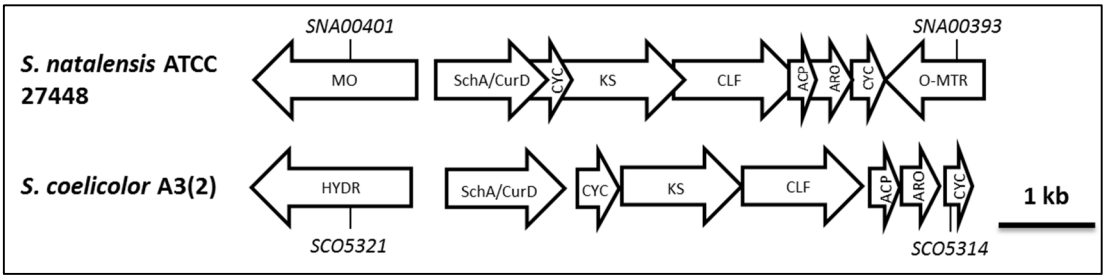

Figure 10 – Organization of the S1C2 putative polyketide subcluster in *S. natalensis* ATTC 27448, together with the *whiE* cluster in *S. coelicor* A3(2). (O-MTR, O-methyltransferase; CYC, cyclase; ARO, aromatase; ACP, acyl carrier protein; CLF, chain length factor; KS, ketosynthase; SchA/SurD, SchA/SurD domain containing protein; MO, monooxygenase; HYDR, hydroxylase) [Adapted from Kelemen *et al.*, 1998].

Assuming that the type II-PKS its responsible for the biosynthesis of the spore-associated pigment and that in other *Streptomyces* strains this cluster do not have an associated terpene sub-cluster (Bentley *et al.*, 2002), we can hypothesise that S1C2 is not a hybrid cluster. Moreover, to the best of our knowledge, a hybrid PKS-II-terpene cluster has not been described in the literature.

Regarding post-PKS encoding genes, *SNA00393* codes for a putative O-methyltransferase, and *SNA00401* codes for a putative monooxygenase, suggesting the introduction of a methyl group or/and a hydroxyl group in the final polyketide molecule.

Regarding the terpene core genes, BLASTp analysis revealed the presence of a 1-deoxy-D-xylulose-5-phosphate synthase (DSX) encoding gene (SNA00410). This enzyme is responsible for the first committed step of the MEP pathway, suggesting that this terpene is produced through this pathway (Fig. 11). Moreover, an IPP delta-isomerase (SNA00427), a

dimethylallyltranstransferase (SNA00414), a phytoene synthase (PSY) (SNA00416), a squalene synthase (SQS) (SNA00417), and a squalene/phytoene dehydrogenase – SNA00415 (identity of 85% (379/445)) – were also identified (Table A2; Fig. 11). The presence of these proteins suggest that the end terpene should have a similar structure as squalene or/and phytoene.

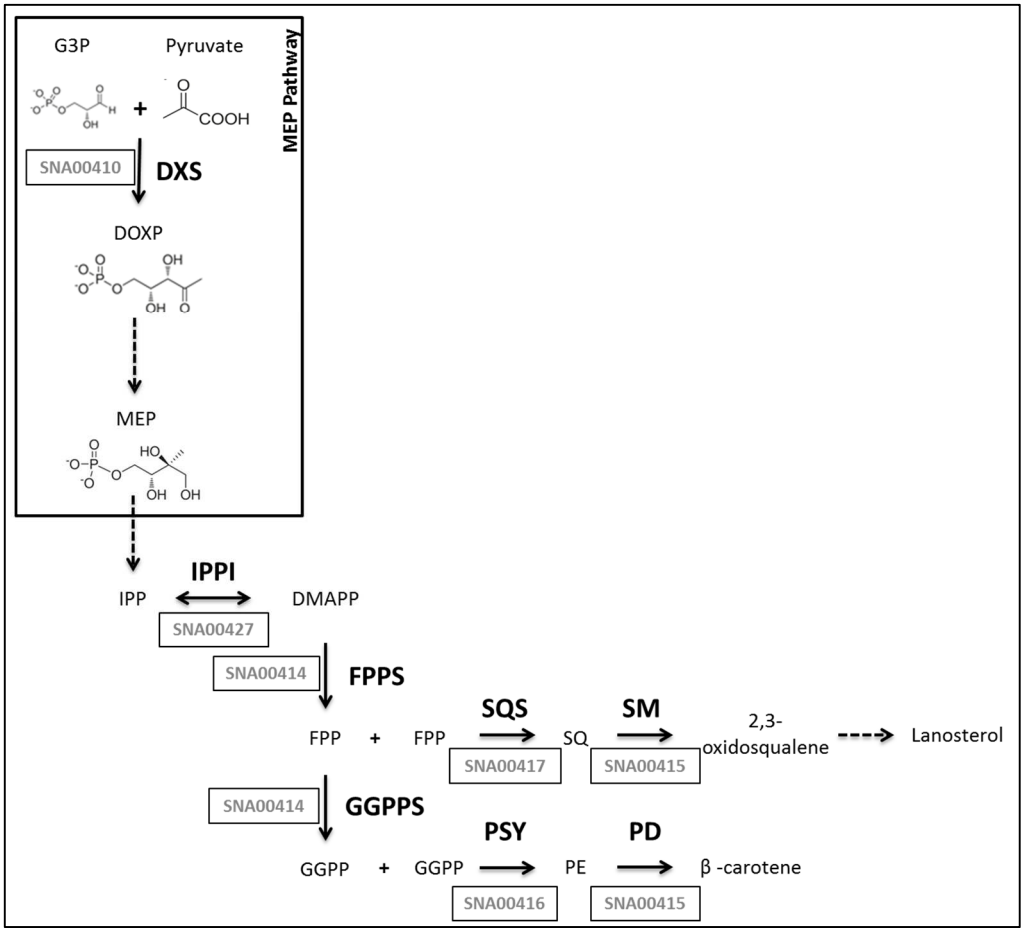

Figure 11 – Schematic representation of the proposed possible biosynthetic pathways for the S1C2 terpene biosynthesis. Once formed by from MEP pathway, IPP is isomerized into DMAPP by an IPP delta-isomerase, coded by SNA00427. Next, dimethylallyltranstransferase, coded by *SNA00414*, converts DMAPP and IPP into FPP (GPP is created in an intermediate step). Then, either (1) FPP can be condensed with another FPP molecule to form SQ or (2) it undergoes the addition of another molecule of IPP to form GPP. If (1), SQS, coded by *SNA00417*, is the enzyme that catalyses the mentioned reaction. Once generated, the SQ is converted into 2,3-oxidosqualene. The latter can then be converted into lanosterol by the lanosterol synthase (an enzyme not identified within S1C2 gene cluster). If (2), the dimethylallyltranstransferase also acts as GGPPS, leading to the formation of GGPP. Two molecules of GGPP can be then condensed by PSY, coded by *SNA00416*, to originate PE. Next, PD converts PE into β-carotene. It may also be possible that the end product results for a combination of both of the later biosynthetic pathways. (G3P, glyceraldehyde-3-phosphate; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; DOXP, 1-deoxy-D-xylulose 5-phosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate; IPP, isopentenyl diphosphate; IPPI, isopentenyl diphosphate isomerase; DMAPP, dimethylallyl diphosphate; FPPS, GGPPS, dimethylallyltranstransferase; FPP, farnesyl diphosphate; SQS, squalene synthase; SQ, squalene; SM, squalene dehydrogenase; GGPP, geranylgeranyl diphosphate; PSY, phytoene synthase; PE, phytoene; PD, phytoene dehydrogenase).

Furthermore, the identification of 11 putative tailoring genes suggests a highly modified end product. A glycosyl transferase (SNA00421), a mannose-1-phosphate guanylyltransferase 1 (SNA00424), a putative dehydrogenase (SNA00423), two aminotransferases (SNA00409 and

SNA00419), a glutamine amidotransferase (SNA00432) and two transferases (SNA00422 and SNA00425) – have been identified as tailoring genes.

Regarding CSR, the antiSMASH analysis retrieved two regulators: an AraC-family member (SNA00399) and a protein assigned as regulatory protein (SNA00428). Manual curation unveiled the presence of three additional putative CSR: an AbaA-like protein (SNA00403), a XRE-like protein (SNA00404) and an AraC-family member (SNA00433) (Table A2). Moreover, *SNA00399* should not encode for a regulator. Instead, as previously mentioned, it should encode for a polyketide cyclase that holds a cupin domain between aa 41-111 (PFAM07883). From the remaining four proteins, one is likely to act positively: SNA00433 (AraC-family member). SNA00404 (coding for a XRE-like protein) is likely to be a negative regulator. Based on protein prediction we were not able to assign a regulatory function to SNA00403 (an AbaA-like protein) and SNA00428. In an attempt to identify the subcluster regulators just by the distance to the core biosynthetic genes, it can be said that SNA00428 and SNA00433 belong to the terpene subcluster. Since *SNA00403* and *SNA00404* are equally close to both subclusters core genes (approximately 3 and 5 kb to the PKS-II and terpene subclusters, respectively) they can regulate either or even both.

As a final remark of S1C2 biosynthetic gene cluster it should be mentioned that it is possible that both PKS-II and terpene sub-clusters may be involved in the production of the same metabolite. However, due to (1) the cluster and subcluster analysis, (2) the distance between them (approximately 8 kb) and (3) the chemical families of the predicted compounds, it is likely that they do not form a hybrid cluster, but are two clusters coding for a type II PKS (*SNA00394-SNA00402*), hereafter named S1C2.1, and a terpene (*SNA00409-SNA00435*) hereafter named S1C2.2. In that case, S1C2 may stand as an example of what could be a weakness of antiSMASH, emphasizing the importance of manual curation of the results presented by the software. Nevertheless, the confirmation of the hybrid nature of the cluster can only be achieved through a functional characterization of the clusters and, more importantly, through the isolation and structure characterization of the final compound.

3.1.2.2 Assessing the transcription of selected genes of S1C2 biosynthetic gene cluster

Cluster S1C2 was assigned by antiSMASH as a hybrid PKS-terpene cluster. For transcription studies, three structural genes, two of S1C2.1 sub-cluster and one of S1C2.2 sub-cluster, have been selected: *SNA00397*, a polyketide chain elongation factor (also referred as a ketosynthase) encoding gene, *SNA00399*, a gene coding for a polyketide cyclase and *SNA00414* that codifies for a putative dimethylallyltranstransferase. Furthermore, the putative regulators

encoding genes, *SNA00403* (AbaA-like protein), *SNA00404* (XRE-family), *SNA00428* and *SNA00433* (AraC-family) have also been selected (Fig. 12).

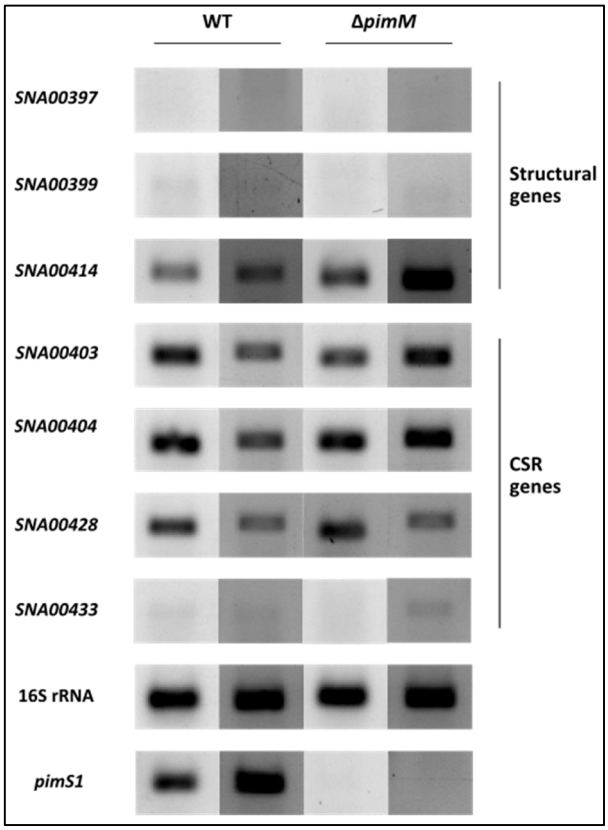

Figure 12 – Transcription analysis of S1C2 selected genes by RT-PCR. Transcripts of *SNA00397* (polyketide chain elongation factor encoding gene), *SNA00399* (AraC-family encoding gene), *SNA00414* (dimethylallyltransferase encoding gene), *SNA00403* (AbaA-like encoding gene), *SNA00404* (XRE-family encoding gene), *SNA00428* (regulatory protein encoding gene), *SNA00433* (AraC-family encoding gene), 16S rRNA (positive control) and *pimS1* (positive control for WT and negative control for Δ*pimM*) were analysed by RT-PCR with 30 cycles on cDNA generated from RNA isolated at early stationary phase. Two biological replicates are shown for each strain.

Overall, no differences between the two strains were observed. Regarding the polyketide structural genes *SNA00397* and *SNA00399* no transcripts were detected in both strains (Fig. 12). This fact is consistent with the *in silico* prediction that this polyketide sub-cluster could be responsible for the production of the grey spore pigment. In fact, it was observed, during the present project, that *S. natalensis* does not sporulate in liquid medium, as it happens for some *Streptomyces* (Yague *et al.*, 2012). Therefore, these transcription results further comprove the mentioned observation.

The terpene structural gene *SNA00414* was expressed in both the wild-type and Δ*pimM* strains. This suggests that the terpene sub-cluster is not silent and, additionally, strengthens the hypothesis that S1C2 is indeed two independent sub-clusters rather than of a hybrid one as predicted by antiSMASH.

Regarding the terpene cluster associated transcription regulator genes, *SNA00428* was transcribed in both strains, while *SNA00433* presented low transcript levels. As a regulator

with no assigned category by *in silico* prediction, the results suggest SNA00428 as a positive CSR that might be directly or indirectly regulating *SNA00414*. In turn, SNA00403 as a predicted positive transcription regulator seems not to be directly involved in the transcription of the studied structural gene. Nevertheless, it may still work as a positive CSR, regulating indirectly the expression of the studied gene or acting on only part of the cluster structural genes.

Lastly, both *SNA00403* and *SNA00404* presented high transcription levels. Since SNA00404 was identified by *in silico* analysis as a negative regulator, and since only the terpene structural gene is being expressed, it is likely that this putative regulator is part of the polyketide sub-cluster. In turn, SNA00403, since its role as a regulator was not foreseen, may either act as a positive regulator in the terpene sub-cluster or as a negative regulator in the polyketide sub-cluster.

3.1.3 S1C3 – a putative type III PK biosynthetic gene cluster

3.1.3.1 *In silico* analysis

As above mentioned, PKS-III are mostly found in plants. In fact, it was proposed that type III PKS were plant specific. However, several studies have shown that they are widely present in bacteria and in fungi (Yu *et al.*, 2012). These PKSs are simple homodimers of ketosynthases that do not use ACP for the carbon chain elongation and act directly on the extender units (acyl-CoAs) to catalyse iterative condensations (Fujii, 2008, Song *et al.*, 2006, Yu *et al.*, 2012).

S1C3 (Fig. 13), was identified by antiSMASH as encoding a type III PKS that presented low synteny with other *Streptomyces*.

BLASTp analysis of the proteins coded by S1C3 cluster (Table A3) retrieved a type III PKS, SNA00841 (89% identity with RppA from *Streptomyces albulus*), and a cytochrome P450, SNA00843 (84% identity with a cytochrome P450 from *Streptomyces albulus*). The co-existence of a PKS-III and a cytochrome P450 encoding genes in the same cluster is a feature observed in other streptomycetes type III PKS biosynthetic gene clusters (Fig. 14) (Funa *et al.*, 2005). The RppA of *Streptomyces griseus*, homologous to SNA00841, was the first bacterial PKS identified as a member of the chalcone synthase (CHS) superfamily and categorized as a PKS-III (Funa *et al.*, 1999). Furthermore, the association of RppA with the production of melanin in *S. griseus* (Funa *et al.*, 2002) together with the similarity between the RppA encoding gene from *S. griseus* with other cluster situated RppA-like enzymes suggested the involvement of these enzymes in the biosynthesis of a wider range of secondary metabolites (Bangera and Thomashow, 1999, Pfeifer *et al.*, 2001, Funa *et al.*, 2002).

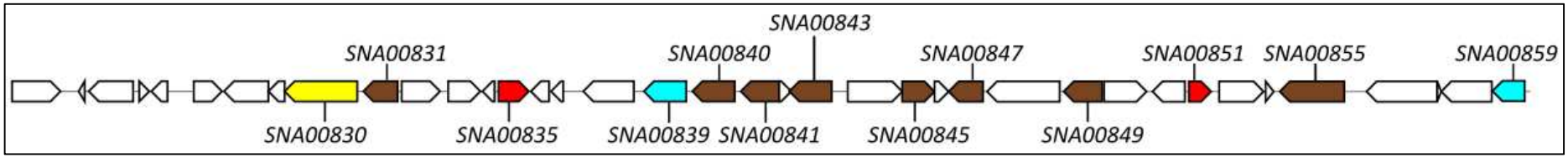

Figure 13 – Schematic representation of cluster S1C3. Brown arrows represent putative structural genes. Blue arrows represent putative transporters encoding genes (*SNA00839* and *SNA00859*). The red arrows represent putative negative CSR encoding genes: *SNA00835* (XRE-family); and *SNA00851* (TetR-family). The yellow arrow (*SNA00830*) represents a putative CSR encoding gene which may either act as activators or repressors: *SNA00830*. White arrows represent other gene.

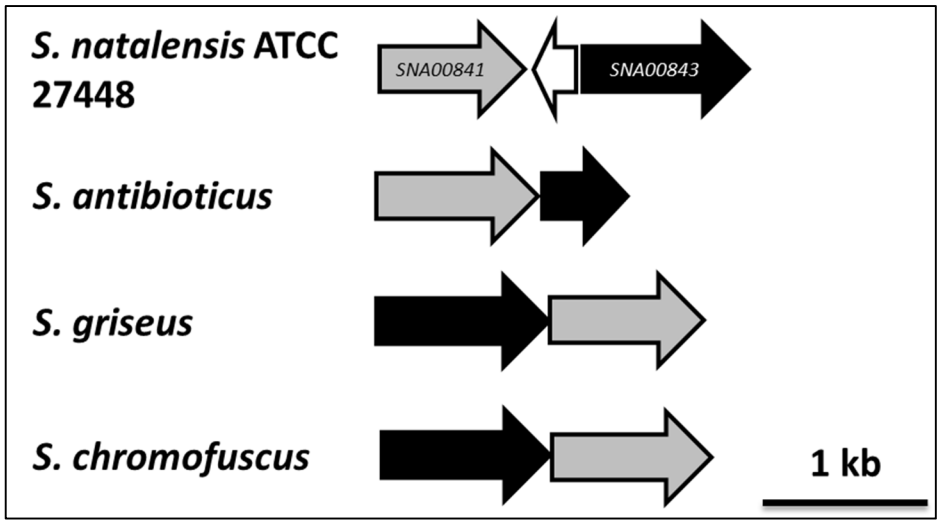

Figure 14 – Schematic representation of some type III PKS gene clusters from streptomycetes. Each arrow represents a different ORF. Grey arrows represent different *rppA* genes coding for type III PKS, while black arrows represent different P-450 encoding genes. The white arrow represents a hypothetical protein [Adapted from Zeng *et al*., 2012].

BLASTp analysis also allowed to identify the putative tailoring genes, which may be responsible for the modification of the compound post-PKS. Two of these genes encode for putative methyltransferases (*SNA00831* and *SNA00845*) which suggests that the polyketide may suffer the addition of two (or more) methyl groups. Moreover, it was found the presence of a gene coding for a putative alcohol dehydrogenase (*SNA00847*) and for a putative acetyltransferase (*SNA00849*). These enzymes catalyse the interconversion between alcohols and aldehydes or ketones and the transference of an acetyl group, respectively. Thus, the end polyketide molecule may contain an extra acetyl group and one of its acyl groups may be converted into a hydroxyl group (OH). Additionally two genes coding for putative transporters have been identified (*SNA00839* and *SNA00859*).

It is also interesting to note that *SNA00841* and *SNA00843* seem to be surrounded by genes somehow involved in the modification or transport of proteins (from *SNA00840* to SNA00859), namely *SNA00840* that encodes for a putative aminopeptidase and *SNA00855* that encodes for a putative S15 family peptidase, among others. Thus, it is likely that those genes that appear to be involved in the peptide metabolism will further functionalize the type III polyketide.

Only two putative regulators were identified, notably SNA00835 (XRE-family) and SNA00851 (TetR-family). SNA00830 was also identified as encoding a putative regulator although no domains were detected in the amino acid sequence that could unveil its role.

3.1.3.2 Assessing the transcription of selected genes of S1C3 biosynthetic gene cluster

S1C3 was identified as a type III PKS encoding cluster. The PKS-III encoding gene *SNA00841* was chosen for transcription analysis together with the putative transcription

regulators encoding genes: *SNA00830*, *SNA00835* and *SNA00851*. Results are displayed in Figure 15.

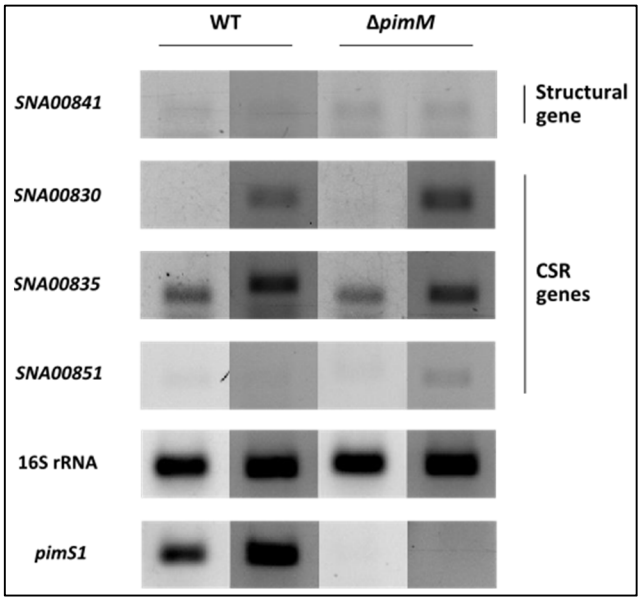

Figure 15 – Transcription analysis of S1C3 selected genes by RT-PCR. Transcripts of *SNA00841* (PKS-III encoding gene), *SNA00830* (regulator encoding gene), *SNA00835* (XRE-family encoding gene), SNA00851 (TetR-family encoding gene), 16S rRNA (positive control) and *pimS1* (positive control for WT strains and negative control for Δ*pimM* strains) were analysed by RT-PCR with 30 cycles on cDNA generated from RNA isolated at early stationary phase. Two biological replicates are shown for each strain.

Altogether, results showed no major differences between strains regarding the expression of S1C3 genes. The structural gene *SNA00841* exhibited low transcription levels in the wild-type and in Δ*pimM*. This way, it was outstood the silent state of the cluster.

SNA00851, encoding a putative TetR-family regulator, presented low transcription levels, similarly to the structural gene. This fact, together with its location and predicted regulatory activity suggests SNA00851 as a positive CSR involved in the control of the PKS-III encoding gene transcription. Although TetR family is mostly composed by transcription repressors (Ramos *et al.*, 2005), there are some documented examples of transcription activators on this family. One of such cases is the *Clostridium tetani* CN655 TetR that was shown to be a positive regulator of the tetanus toxin gene (*tetX*) (Marvaud *et al.*, 1998). The XRE-family member *SNA00835* was transcribed in both strains, which suggests that it acts as a repressor of this gene cluster. Interestingly, this fact is consistent with the prediction of the present gene be coding for a negative regulator. Lastly, the inconsistent expression pattern of *SNA0830* in biological replicates does not allow any prediction regarding its function.

3.1.4 S1C4 – a putative bacteriocin biosynthetic gene cluster

3.1.4.1 *In silico* analysis

Bacteriocins are peptides produced by many bacteria with the purpose of inhibiting the growth of similar or closely related bacterial strains, which may exercise competition in milieu

(Karpiński and Szkaradkiewicz, 2013, Nigam *et al.*, 2014). These compounds are widely used as food preservative (Cleveland *et al.*, 2001) and in the clinical practice against bacterial pathogens, cancer, and anti-inflammatory diseases (Karpiński and Szkaradkiewicz, 2013).

A relatively short precursor peptide is first translated and then modified. This modification includes at least one proteolytic step that cleaves the precursor peptide on its N-terminal extension, at specific processing sites, giving rise to a smaller and active fragment (Fig. 16). The N-terminal leader sequence presumably facilitates interaction with the transporter and/or keeps the peptide inactive until it has been secreted from the cell. Further post-translational modifications, such as macrocyclization or formylation, among others, may also take place. The end bacteriocin must be then secreted, by the action of transporters. In some cases, as it happens in nisin biosynthesis, the export is tightly coupled to proteolysis (McIntosh *et al.*, 2009). As a final note regarding bacteriocin biosynthesis, the genes encoding precursor peptides often cluster with genes encoding modifying enzymes, secretion, and also immunity/resistance genes.

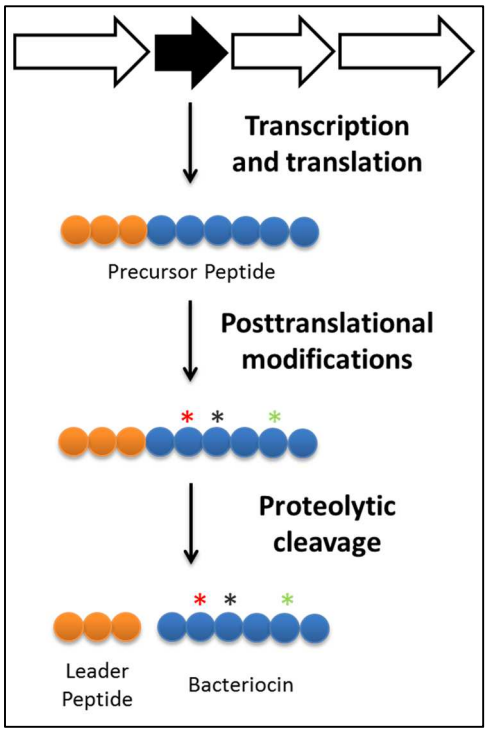

Figure 16 – Bacteriocins biosynthesis. The gene encoding for the precursor peptide, usually clustered with genes encoding modifying enzymes, secretion, and also immunity genes, is initially transcribed and then translated. Then, posttranslational modifications take place, including a proteolytic cleavage allows the separation of the leader peptide from the active peptide [Adapted from McIntosh *et al*., 2009].

S1C4 (Fig. 17), a cluster predicted to be responsible for the biosynthesis of a bacteriocin, displayed low synteny with homologous gene clusters presented by antiSMASH. Analysis of the 25 genes that are part of S1C4 (Table A4) showed that two may encode the bacteriocin

precursor peptide: *SNA01164* and *SNA01165*. Despite the fact that BLASTp retrieved homologies with hypothetical proteins (Table A4) protein domain analysis showed that both proteins have a TIGR04222 domain (from aa 1 to aa 241 and from aa 7 to aa 254, respectively). The majority of the proteins with a TIGR04222 domain as described by this model have a C-terminal sequence that consists of extremely low-complexity sequence, rich in Ser or in Gly interspersed with Cys that is present in SNA01164. That C-terminal region resembles ribosomal natural product precursors, although there is no evidence that C-terminal regions of these proteins undergo any modification or have any such function (Marchler-Bauer *et al.*, 2013). Furthermore, *SNA01166* encodes a putative tripeptidyl aminopeptidase (TAP) (71% identity). TAP is a member of the S33 peptidase family, a subgroup of serine proteases with amino proline peptidase activity (Rawlings and Barrett, 1994). Thus, it can be hypothesized that *SNA01166* encodes the proteolytic enzyme responsible for the cleavage of the peptide precursor.

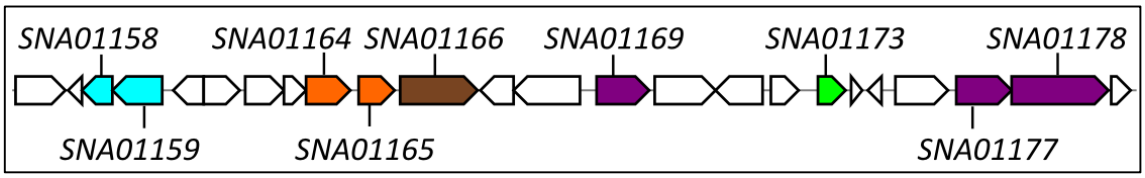

Figure 17 – Schematic representation of cluster S1C4. The brown arrow represents *SNA01166* that encodes a putative tripeptidyl aminopeptidase that might work as the proteolytic enzyme responsible for the cleavage of the peptide precursor. The orange arrows represent SNA01164 and SNA01165, two genes encoding for putative hypothetical proteins that hold a TIGR04222 domain. These domains are typical in proteins with a C-terminal region that resembles ribosomal natural product precursors. Purple arrows represent genes encoding for enzymes related with fatty acids metabolism. Blue arrows represent putative transporters encoding genes. The green arrow represents *SNA01173* that encodes a putative positive CSR (LuxR-family). White arrows represent other genes.

Three genes appear to be involved in degradative pathways of fatty acids: *SNA01169*, that codes for a putative beta-ketoadipyl CoA thiolase, *SNA01177*, which encodes a putative acetyl-CoA acetyltransferase and *SNA01178*, which encodes a putative 3-hydroxyacyl-CoA dehydrogenase. It is currently known that the addition of isoprene-derived subunits to a core molecule (prenylation) is a post translational reaction extremely common in the natural products world. Since the final products of fatty acids degradation are precursors for the biosynthesis of terpenes the presence of the mentioned three genes suggests that isoprene precursors may further be added to the bacteriocin.

Two genes coding for ABC transporters (*SNA01158* and *SNA01159*) were also identified. These proteins could be involved in the secretion of the bacteriocin synthesized by S1C4.

Lastly, it has been determined that S1C4 contains a putative positive regulator: SNA01173. This gene has been identified as coding for a LuxR-family member. Moreover, the protein domain analysis revealed the presence of a signal receiver domain, that receives the

signal from the sensor partner in a two-component systems (aa 1-112) and a C-terminal DNA-binding domain of LuxR-like proteins (aa 138-194).

3.1.4.2 Assessing the transcription of selected genes of S1C4 biosynthetic gene cluster

S1C4 was identified by antiSMASH as a bacteriocin biosynthetic gene cluster. From all its twenty five genes, as none clearly identified as a prepeptide encoding gene, *SNA01166*, that encodes a putative TAP, was believed to be the most likely to be related with bacteriocins biosynthesis, which justified its selection for RT-PCR experiments. Concerning to transcription regulators, only *SNA01173* (LuxR-family) was identified and hence subjected to the transcription analysis.

The transcription of both genes was residual in both strains (Fig. 18), suggesting that this cluster was silent in the wild-type and in Δ*pimM* strains. This result is coherent with the bioinformatics prediction of S1C4 being a bacteriocin type gene cluster, since these metabolites are produced during the primary phase of growth and mycelium samples were harvested at the beginning of the stationary phase (Beasley and Saris, 2004). In addition, it can be rightly argued that the repressed state of *SNA01173*, together with its genomic location and predicted *in silico* characterization, indicates that it acts as the positive regulator of this cluster. Therefore, *SNA01173* was selected for overexpression towards S1C4 activation.

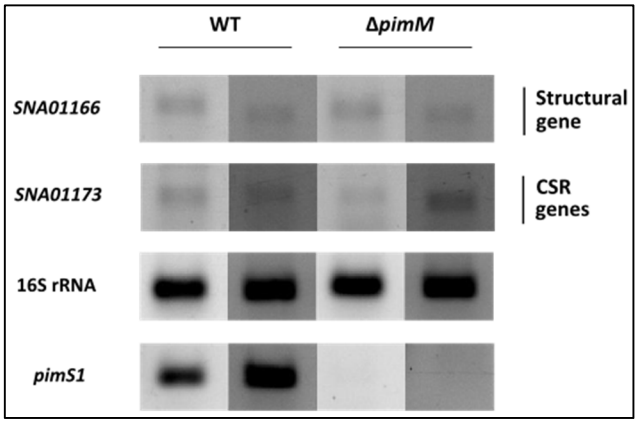

Figure 18 – Transcription analysis of S1C4 selected genes by RT-PCR. Transcripts of *SNA01166* (TAP encoding gene), *SNA01173* (LuxR-family encoding gene), 16S rRNA (positive control) and *pimS1* (positive control for WT strains and negative control for Δ*pimM* strains) were analysed by RT-PCR with 30 cycles on cDNA generated from RNA isolated at early stationary phase. Two biological replicates are shown for each strain.

3.1.5 S1C5 – a putative NRP biosynthetic gene cluster

3.1.5.1 *In silico* analysis

Nonribosomal peptide synthetases (NRPS), are multimodular and multifunctional enzymes, known for their ability to produce bioactive compounds using amino acids in both the L- and D- forms as precursors (Hwang *et al.*, 2014). NRPS are composed of distinct modules, each one responsible for the incorporation of one specific amino acid into the final

product. By their turn, modules are subdivided into catalytically independent domains, namely adenylation (A), peptidyl carrier protein (PCP or thiolation) and condensation (C). The A domain is responsible for the recognition and consequent activation of the amino acid building block by the formation of an amino acyl adenylate intermediate. Then, the activated substrate is transferred to the PCP where it is covalently bound through a thioester linkage, leading to the activation of this domain. Finally, C domain is responsible for the peptide bond formation with the downstream amino acyl unit that is tethered to the PCP of the adjacent module (Sieber and Marahiel, 2005). The last module presents a thioesterase (TE) domain that terminates the elongation process, releasing the molecule. In some cases it can also be found epimerization domains (E), which are responsible for the conversion of the PCP tethered aminoacyl substrate from the L to the D form (Fischbach and Walsh, 2006, Rix *et al.*, 2002, Walsh and Fischbach, 2010).

S1C5 (Fig. 19) was retrieved by antiSMASH as a putative nonribosomal peptide encoding biosynthetic gene cluster. Two large NRPS encoding genes were identified: *SNA01225* (8148 bp) and *SNA01226* (18966 bp) (Table A5).

The first, SNA01225 displayed two modules, predicted to incorporate two amino acids. In turn, SNA01226 has five modules. Therefore it is predicted to incorporate five amino acids.

AntiSMASH amino acid prediction based on the A domains retrieved that the module 3 and 4 from SNA01226 should introduce an alanine into the end product, while SNA01225 first module should incorporate a valine (Fig. 20). SNA01225 sequence displayed a TE domain at the end which may indicate the end of the compound biosynthesis. However three additional NRPS domain encoding genes were also identified from the protein domain analysis: SNA01211 (PCP); SNA01212 (A domain); SNA01213 (TE domain). These three domains may function as an additional NRPS module predicted to introduce a threonine residue.

Worth to note that if the TE domain present in SNA01213 is actually part of the module constituted by SNA01211 (PCP) and SNA01212 (A domain) there will be two TE domains in the NRPS. To the best of our knowledge, until this date no NRPS system has been described as holding two TE domains. Alternatively, SNA01211 and SNA01212 could constitute a loading module responsible for feeding SNA01226 with a shorter unit which could be a threonine.

The S1C5 ClusterBlast analysis showed a high level of synteny with the biosynthetic gene cluster of himastatin from *Streptomyces himastatinicus* ATCC 53653 (Fig. 21). Particularly the genomic region that includes genes *SNA01215* to *SNA01224*.

Himastatin is synthethized by a 7-module NRPS and its structure harbours an unusual symmetric cyclohexadepsipeptide dimer and a peptidyl backbone (Fig. 22).

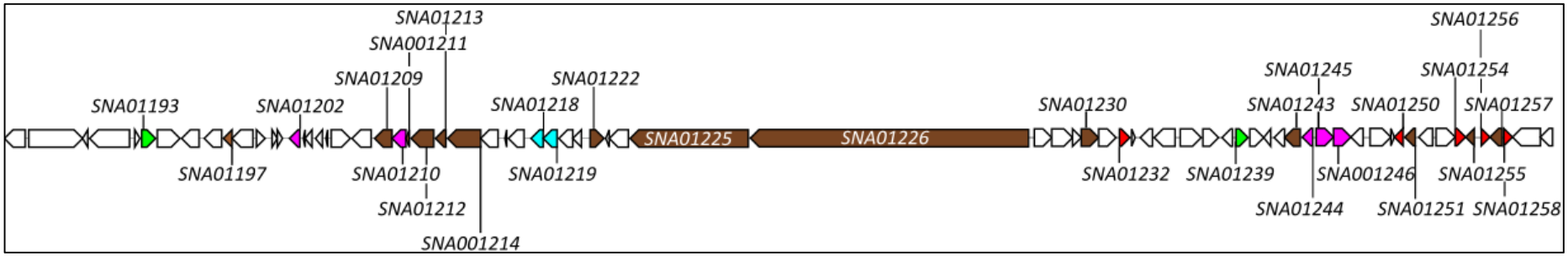

Figure 19 – Schematic representation of cluster S1C5. The brown arrows represent the structural genes, including the core NRPS synthases encoding genes *SNA01211*, *SNA01212*, *SNA01213*, *SNA01225* and *SNA01226*. Pink arrows represent peptide modifying enzymes encoding genes that may or may not be involved in the NRP tailoring. Blue arrows represent putative transporters encoding genes: *SNA01218* and *SNA01219*. Green arrows represent *SNA01193* (RNA polymerase sigma factor) and *SNA01239* (MerR-family) that encode for putative positive CSRs, while red arrows represent putative negative CSRs: *SNA01232*, *SNA01256* and *SNA01258* (TetR-family); *SNA01250* (DeoR-family); and *SNA01254* (PaaX-family). White arrows represent other genes.

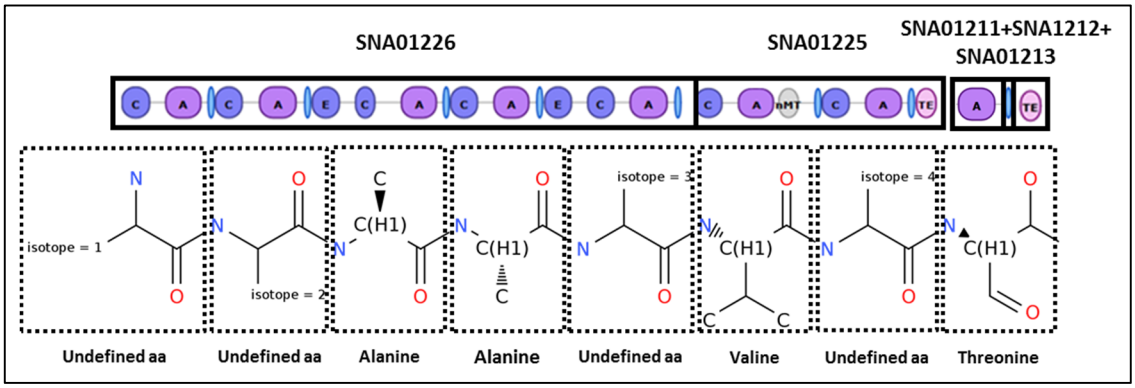

Figure 20 – AntiSMASH prediction for protein domain specificities and chemical structure of the S1C5 NRP. SNA01226 A domain specificities suggest the incorporation of two alanines by the 3rd and 4th modules. In turn, SNA01225 A domain specificities suggest the incorporation of a valine by the 1st module. The module putatively formed by SNA01211 (PCP); SNA01212 (A domain); SNA01213 (TE domain) was predicted to incorporate a threonine. Modules with no assigned specific aminoacid hold isotope as residue. (A, adenylation domain; C, condensation domain; E, epimerization domain; nMT, N-methyltransferase domain; TE, thiosterase domain; the PCP domains are represented by light blue balls).

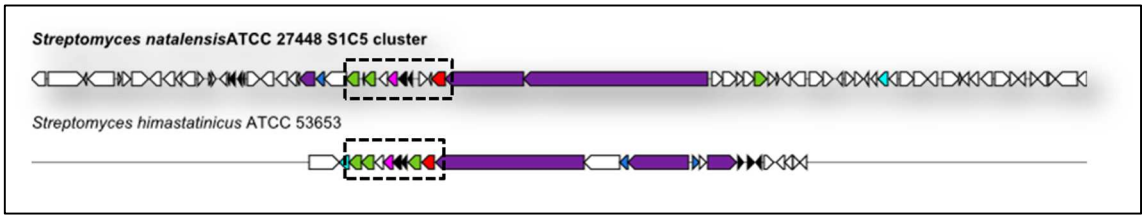

Figure 21 – Top S1C5 homologous gene cluster: himastatin biosynthetic gene cluster from *Streptomyces himastatinicus* ATCC 53653 (accession number FR823394.1_c1). Homologous genes are shown in identical colours, whereas white-coloured genes have no Blast hits between the gene clusters. The most conserved region between the two clusters is marked within a dashed box and includes three P450-like enzymes, one peptide monooxygenase and one methyltransferase encoding genes (green, red and light blue arrows, respectively). Other identical genes are represented by black arrows.

It is a fact there are similarities between both himastatin biosynthetic gene cluster from *Streptomyces himastatinicus* ATCC 53653 and S1C5. Nevertheless, despite the fact that the end NRP produced from S1C5 could share resemblances with himastatin, it is clear that they will be to different compounds. Moreover, the differences between the two NRP molecules may then be translated in interesting alterations regarding their bioactivity.

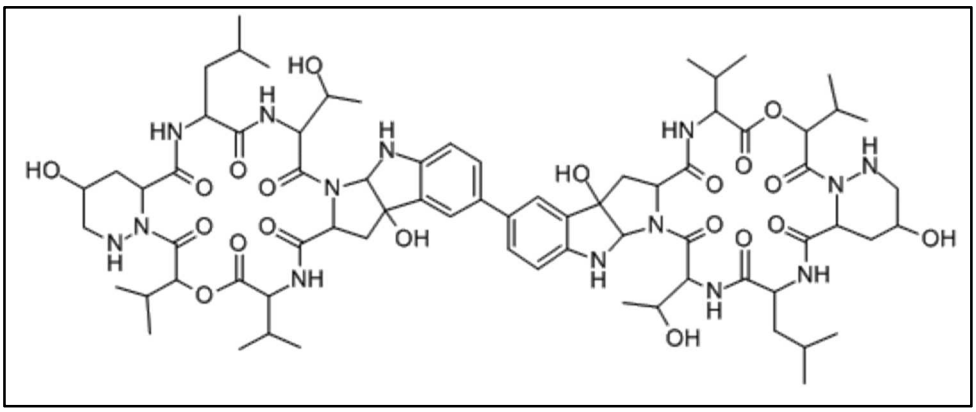

Figure 22 – Himastatin chemical structure [Adapted from Zhang *et al.*, 2013].

Regarding putative tailoring enzymes, SNA01226 owns itself one methyltransferase domain (from aa 941 to aa 1159). Additionally, two other putative methyltransferases encoding genes were also detected (SNA01243 and SNA01247). Thus, it is likely that the NRP undergoes three methylations, of which one should take place at the valine that is predicted to incorporate the growing chain (module 1 of SNA01225). Moreover, three P450-like enzymes (SNA01215, SNA01216 and SNA01231), two acetyltransferases (SNA01197 and SNA01222), two hydroxylases (SNA01251 and SNA01255), one alpha/beta hydrolase fold protein (SNA01209), one peptide monooxygenase (SNA01224), one aminotransferase (SNA01230) and one dehydrogenase (SNA01257) were identified.

Worth to note that several putative peptide modifying enzymes related have been detected: SNA01202 (amidinotransferase); SNA01210 (chlorinating dioxygenases); SNA01214 (peptidase); SNA01244 (biopterin-dependent aromatic amino acid hydroxylase family protein); and SNA01245 (pyridoxal-dependent amino acid decarboxylase). Thus, the NRP may further undergo through a functionalization performed by those enzymes typically involved in the peptide metabolism.

Two ABC transporters (SNA01218 and SNA01219) were also detected. Therefore, it is possible that both SNA01218 and SNA01219 are part of the S1C5 NRP biosynthetic pathway as encoding proteins responsible for excretion of the end peptide.

Regarding the CSRs present in this cluster, two putative transcription activators were identified: SNA01193 (an RNA polymerase sigma factor) and SNA01239 (a MerR-like protein). In addition, five genes encoding for putative negative regulators, three from TetR-like (SNA01232, SNA01256 and SNA01258), one belonging to the DeoR-family (SNA01250) and one of the PaaX-family (SNA01254) were also detected. It should be noted that SNA01193 is located approximately 20 kb from the biosynthetic core genes and most likely is not part of this cluster.

3.1.5.2 Assessing the transcription of selected genes of S1C5 biosynthetic gene cluster

For S1C5 transcription analysis, seven genes were selected: the putative transcription regulators *SNA01239*, *SNA01239*, *SNA01250*, *SNA01254*, *SNA01256* and SNA01258 and a structural gene putatively encoding for a NRPS: *SNA01225*.

As shown in Figure 23, the structural gene *SNA01225* is poorly transcribed in both the wild-type and Δ*pimM*, suggesting that the transcription of S1C10 biosynthetic gene cluster is silenced in both strains.

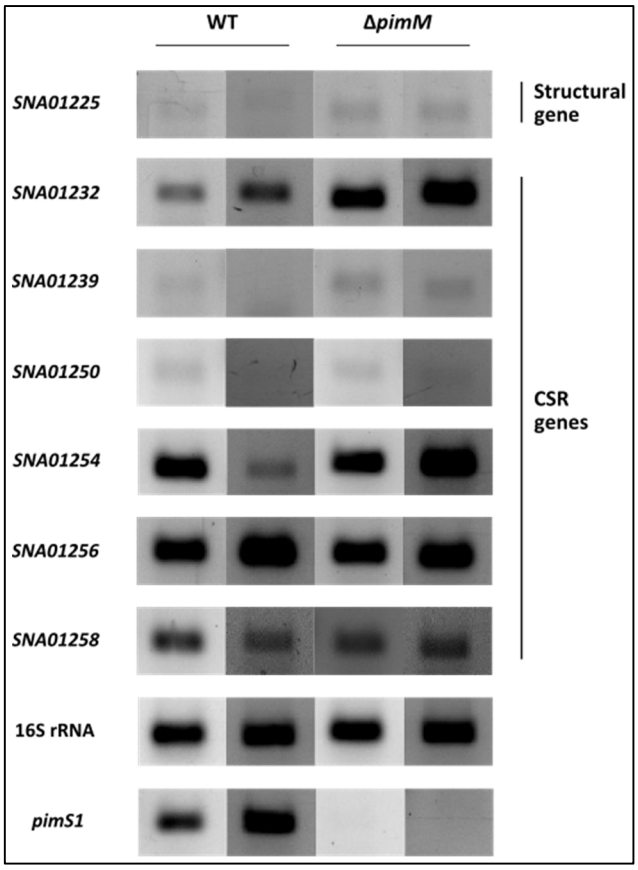

Figure 23 – Transcription analysis of S1C5 selected genes by RT-PCR. Transcripts of *SNA01225* (NRPS encoding gene), *SNA01232* (TetR-family encoding gene), *SNA01239* (MerR-family encoding gene), *SNA01250* (DeoR-family encoding gene), *SNA01254* (PaaX-family encoding gene), *SNA01256* (TetR-family encoding gene), *SNA01258* (TetR-family encoding gene), 16S rRNA (positive control) and *pimS1* (positive control for WT and negative control for Δ*pimM*) were analysed by RT-PCR with 30 cycles on cDNA generated from RNA isolated at early stationary phase. Two biological replicates are shown for each strain.

Regarding the results for the CSRs present in this cluster, *SNA01239*, similarly to the structural gene, also presented low transcript levels. This result, together with its *in silico* caractherization and genomic location, suggests that it may act as positive regulator. This way, *SNA01239* overexpression stands as a promising strategy to awake this cluster and identify the final compound.

With the exception of *SNA01250*, all the remaining genes encoding putative CSRs seem to have a high level of transcripts. Consequently, owing to their genomic location within S1C5 and its *in silico* prediction, this protein might be involved in the repression of the gene cluster. Regarding the DeoR-family regulator encoding gene *SNA01250*, its repressed expression state may indicate that it acts as positive regulator. Such activity would not be expectable since DeoR-like regulators are usually repressors (Skerlová *et al.*, 2014). However there are reported DeoR-family members that own a transcription activator role. The *fruR* gene from *Spiroplasma citri* that codes for a DeoR-family regulator activates the transcription of the fructose operon

(Gaurivaud *et al.*, 2000, Gaurivaud *et al.*, 2001) and stands as an example of the previously mentioned.

It should also be mentioned that *SNA01232* appears to have higher transcription levels in the Δ*pimM* strain rather than in the wild-type. This result, together with the fact that SNA01232 was predicted as a putative negative CSR according to bioinformatics analysis, may indicate that this protein is indeed repressing the activation of the structural gene.

3.1.6 S5C20 – a putative type I PK-NRP biosynthetic gene cluster

3.1.6.1 *In silico* analysis

Polyketides and nonribosomal peptides, although different in overall structure and fundamental chemical building blocks, exhibit striking similarities in their biosynthetic assembly mechanisms. These similarities allow the formation of hybrid clusters that contain elements of each class (Du *et al.*, 2001, Fisch, 2013).

The combination of PKS and NRPS has already been described. In these systems, PKS and NRPS modules are combined in a hybrid modular system (Fisch, 2013). For example, in the antibiotic bacillaene biosynthesis, BaeJ protein combines three PKS with a NRPS module which incorporates glycine (Chen *et al.*, 2006).

S5C20 (Fig. 24) was identified as a hybrid PKS-NRPS biosynthetic gene cluster that presented low synteny with PKS-NRPS hybrid clusters. According to the presented functional annotation of its different domains, SNA06257 holds one A and one PCP domains, typical of NRPS system, and one PKS module comprising a KS-AT-DH-KR-ACP architecture (Fig. 25). Moreover, *SNA06258* was identified as a TE encoding gene that could be responsible for the end of the elongation process.

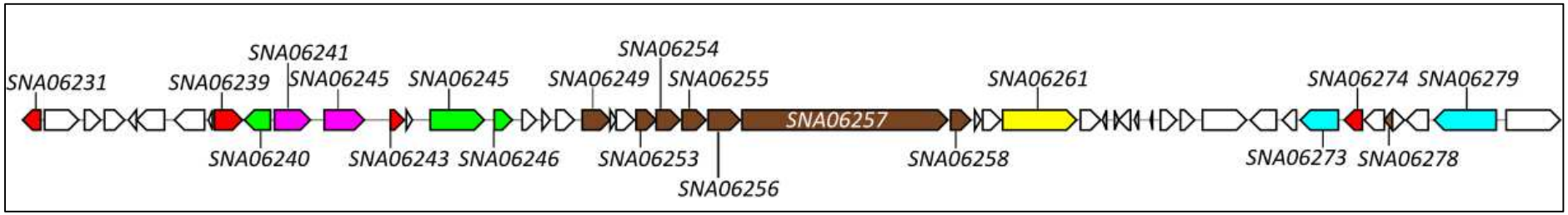

Figure 24 – Schematic representation of cluster S5C20. The brown arrows represent the structural genes, such us the putative mixed PKS-NRPS encoding gene. Pink arrows represent peptide modifying enzymes encoding genes that may or may not be involved in the hybrid PK-NRP tailoring. Blue arrows represent putative transporter genes: *SNA06273* and *SNA06279* (both putative DHA2-family drug resistance MFS transporters). Green arrows represent all the three putative positive CSR encoding genes: *SNA06240* (LysR-family); *SNA06245* (AfsR-family); *SNA06246* (LuxR-family). Red arrows represent putative negative CSRs encoding genes: *SNA06231* (PadR-family); *SNA06239* (GntR-family); *SNA06243* (AsnC-family); and *SNA06275* (TetR-family). Yellow arrow represent a gene encoding for a transcription regulator that might act as repressor or activator: *SNA06261* (serine/threonine protein kinase). White arrows represent other genes.

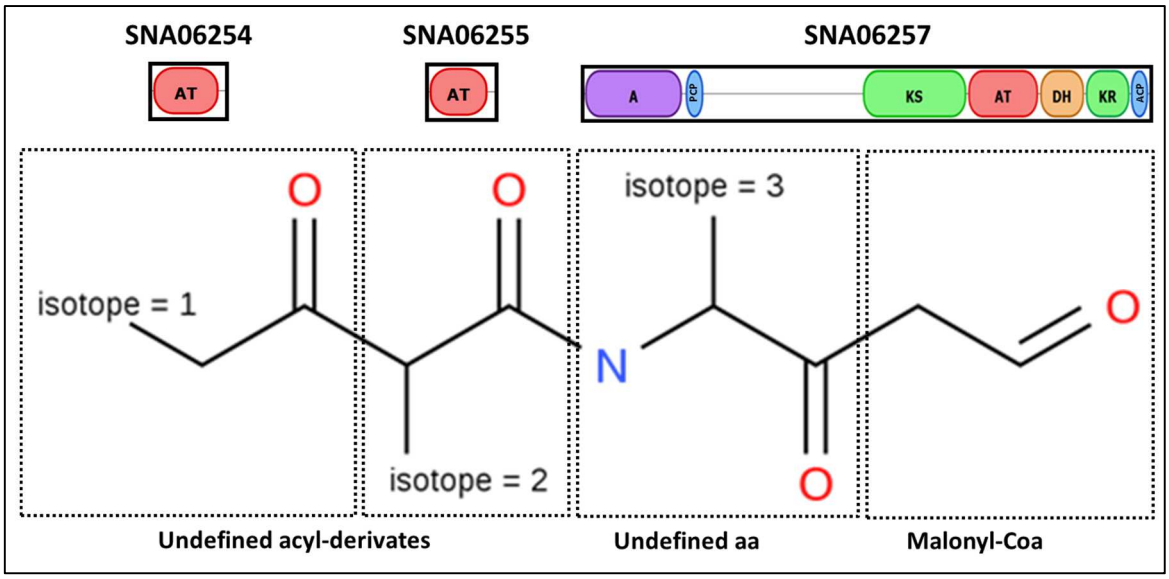

Figure 25 – AntiSMASH prediction for protein domain specificities and chemical structure of the S5C20 hybrid PK-NRP. AntiSMASH substrate specificities predictions of PKS and NRPS modules are based on the active sites of their corresponding AT and A domains. Thus, each AT and the A domain have an acyl-derivate or amino acid associated, respectively, even if they do not have the remaining domains required for a minimal PKS or NRPS. In the particular case of the NRPS, it is likely that the module may work as a loading one. Therefore, it does not require the lacking C domain at fault (AT, acyl transferase; A, adenylation domain; PCP, peptide carrier protein; KS, ketosynthase; DH, dehydrogenase; KR, ketoreductase; ACP, acyl carrier protein). antiSMASH prediction for protein domain specificities of S5C20. antiSMASH prediction for the core chemical structure of the final compound produced by S5C20.

AntiSMASH amino acid and acyl units prediction based on the A and AT domains (for NRPS and PKS, respectively) retrieved that SNA06257 NRPS module introduces an unspecific amino acid, while the PKS module introduces a malonyl unit. Since the protein domain analysis showed that the NRPS module lacks a C domain required for a minimal NRPS, the NRPS domains should act as a loading module

In addition to the molecular entities introduced by SNA06257, antiSMASH retrieved that two more acyl derivative units should also be incorporated into the hybrid PK-NRP molecule by two putative independent PKS modules. These modules include an AT domain suggested to be present in SNA06254 and in SNA06255. However, a minimal PKS requires, besides the AT domain, one ACP and one KS domains. Since it was not detected any KS and only one putative ACP (SNA06250) was identified in the putative mixed PKS-NRPS surrounding genes, it cannot be ascertain that the end compound will undergo through the addition of these two acyl derivative units.

Concerning tailoring genes, one putative acetyltransferase has been detected (*SNA06278*) which indicates that the end product might have an extra acetyl group, and also several encoding proteins that catalyse typical reactions involving amino acids, namely a putative aminotransferase (*SNA06241*), a lysine 2,3-aminomutase (*SNA06242*) and a glycine

amidinotransferase (*SNA06256*). Thus, these enzymes involved in the peptide metabolism may further functionalize the hybrid PK-NRP.

BLASTp analysis also retrieved two putative transporters, both identified as Drug:H^+ Antiporter family 2 (DHA2) (SNA06273 and SNA06279). They are transporters of the Major Facilitator Superfamily (MFS) involved in multidrug resistance (MDR). The DHA2 family members were found to participate in the export of structurally and functionally unrelated compounds or to be involved in the uptake of amino acids (Dias and Sa-Correia, 2013). Thus, it is possible that both SNA06273 and SNA06279 are involved in the biosynthetic pathway of this hybrid compound either for its secretion or for the uptake of amino acids required for its biosynthesis.

Regarding putative CSR, eight putative transcription regulators were detected, out of which three are likely to be transcription activators: SNA06240 (LysR-family), SNA06245 (AfsR-family) and SNA06246 (LuxR-family). The remaining regulators, three are regulators whose function is predicted to be negative: SNA06239 (GntR-family), SNA06243 (AsnC-family), and SNA06274 (TetR-family). Furthermore, SNA06261 was identified as a serine/threonine protein kinase.

SNA06231 and *SNA06274* also code for putative regulatory proteins. However they were assumed to be out of the boundaries of S5C20 biosynthetic gene cluster. The relative distance to the core biosynthetic genes (20 kb and 13 kb, for *SNA06231* and *SNA06274*, respectively) and the fact that these genes are surrounded by genes coding for putative hypothetical proteins or proteins not related with the production neither of a PK nor a NRP support this exclusion.

3.1.6.2 Assessing the transcription of selected genes of S5C20 biosynthetic gene cluster

The six putative regulatory encoding genes *SNA06239*, *SNA06240*, *SNA06243*, *SNA06244*, *SNA06246* and *SNA06261* from S5C20 were selected for transcription analysis together with the structural gene *SNA06257* that codes for a putative mixed PKS-NRPS.

Obtained results showed that *SNA06257* was being poorly transcribed (Fig. 26) in both tested strains, which suggests that S5C20 is a silent cluster.

Regarding the regulatory genes initially proposed as putative positive regulators (*SNA06240*, *SNA06245* and *SNA06246*) only the LuxR-family SNA06246 presented low transcript levels, similarly to *SNA06257*. Therefore, it is likely to be responsible for the activation of S5C20 gene cluster and, consequently, it is the best candidate to be overexpressed. In turn, both *SNA06240* and *SNA06245* showed high transcription levels, especially in Δ*pimM.*

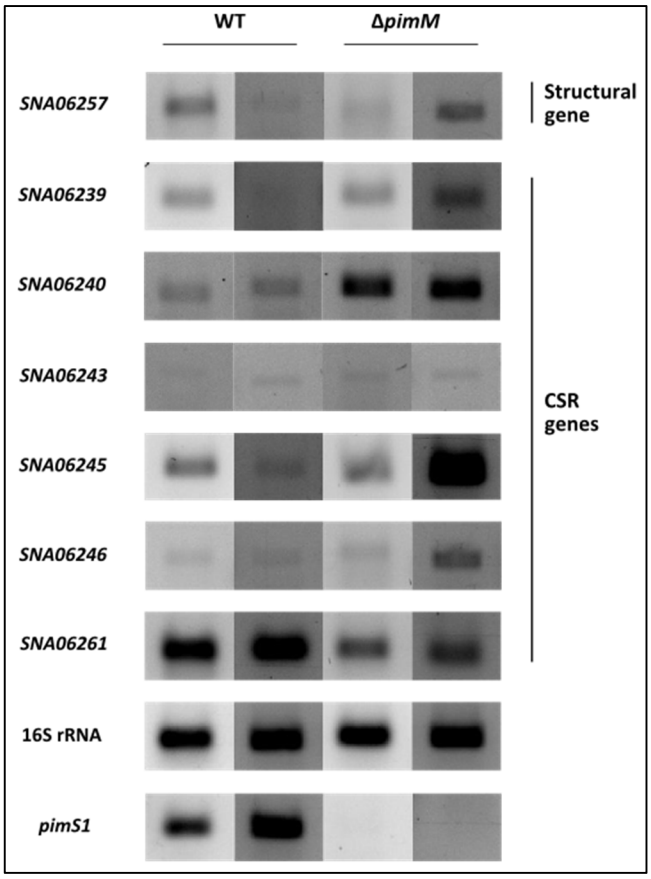

Figure 26 – Transcripton analysis of S5C20 selected genes by RT-PCR. Transcripts of *SNA06257* (mixed PKS-NRPS encoding gene), *SNA06239* (GntR-family encoding gene), *SNA06240* (LysR-family encoding gene), *SNA06243* (AsnC-family encoding gene), *SNA06245* (AfsR-family encoding gene), *SNA06246* (LuxR-family encoding gene), *SNA06261* (serine-threonine kinase encoding gene), 16S rRNA (positive control) and *pimS1* (positive control for WT and negative control for Δ*pimM*) were analysed by RT-PCR with 30 cycles on cDNA generated from RNA isolated at early stationary phase. Two biological replicates are shown for each strain.

A curious trait in this cluster is the presence of genes encoding an AfsR-family regulator (*SNA06245*) and a serine-threonine kinase (*SNA06261*). In *S. colicolor*, AfsR (founding member of the AfsR-family) is a pleiotropic positive regulator involved in the production of secondary metabolites that is activated through the phosphorylation of specific residues catalysed by different kinases (Lee *et al.*, 2002). Thus, it is possible that such a regulatory mechanism may be present in S5C20 cluster, making the overexpression of such genes, either alone or combined, a potential successful strategy for S5C20 activation. Finally, concerning the regulators *SNA06239* (GntR-family) and *SNA06243* (AsnC-family), both presented basal expression levels in both strains. These profiles suggest that both transcription regulators may act as transcription activators of the cluster, instead of negative CSR as predicted in the *in silico* analysis. It should be noted that GntR-family members are usually transcription repressors (Hillerich and Westpheling, 2006). Nevertheless, there are some reported cases for members of this family that act as positive regulators, e.g. PigT, which activates the transcription of the *pigA–O* operon in *Serratia* sp. ATCC 39006 (39006) (Fineran *et al.*, 2005).

3.1.7 Conclusions

The obtained results for the *in silico* analysis stress the need for manual curation of antiSMASH results. Only then it is possible to identify all the putative genes that are part of each biosynthetic gene cluster and to predict what is the role of the proteins that they encode for. Moreover, it is possible to validate or not antiSMASH characterization. For instance, S1C2 is not a hybrid type II PK-terpene biosynthetic gene cluster, but two independent gene clusters, one coding for a type II PK (S1C2.1) and the other coding for a terpene (S1C2.2).

It was also evident in the present analysis the difficulty to predict the regulatory functions of putative CSR. Although some of them are mainly related to a given function (either acting as activators or repressors) it is not guaranteed that they will behave that way.

Furthermore, it was also difficult to obtain a precise prediction of the chemical structure of the end compound, although some shared high similarity and synteny with homologous gene clusters. As mentioned, putative tailoring genes have been identified within all presented biosynthetic gene cluster. Enzymes encoded by such genes should be able to functionalize the specialized metabolite. Nonetheless, S1C2.1 as holding a high synteny with the polyketide spore pigment biosynthetic gene cluster is likely to be responsible for the production of this compound in *S. natalensis*.

Following the antiSMASH data curation and analysis, it was also important to assess the "silent" nature of the clusters. For that purpose the transcription of structural encoding genes (genes that code for proteins responsible for the assembly of the metabolite structural backbone) and putative transcription regulator(s) was determined by RT-PCR. The expression of the selected clusters was assessed not only in *S. natalensis* ATCC 27448 (wild-type), but also in *S. natalensis* Δ*pimM*, a strain that lacks pimaricin production (Anton *et al.*, 2007). Worth noting that although not producing pimaricin, the precursors required for the synthesis of this metabolite should be available.

From all the biosynthetic gene clusters submitted for transcriptional analysis, due to the time constraints, only three could be selected for the remaining tasks. S1C1 and S1C2.2, as they appeared to be not silenced, were not selected for overexpression. S1C2.1 although silenced was predicted to encode for the grey polyketide spore pigment which is already a widely studied product. Therefore, S1C2.1 was not selected for the remaining tasks. Leaving only four clusters, three were chosen to proceed with the project: S1C4, S1C5 and S5C20. The purpose of obtaining higher metabolic diversity was on the basis of this selection.

3.2 Ativating *S. natalensis* gene clusters

As mentioned previously, manipulation of CSRs, either by overexpressing activators or deleting repressors, has proven to be a good strategy for increasing titres of secondary metabolite biosynthesis (Gottelt *et al.*, 2010, Laureti *et al.*, 2011, Scheffler *et al.*, 2013). However, due to the time constraints, it was aimed to validate this activation strategy by overexpressing one typical positive regulator, since gene deletion in *Streptomyces* is still a time-consuming process.

From cluster S1C4, *SNA01173* that codes for a putative LuxR-family transcriptional regulator, has presented low transcription levels, in the same way as the studied structural gene. These results together with the fact that it is the only detected putative CSR justify why *SNA01173* was selected to be overexpressed.

Regarding the cluster S1C5, *SNA01239* (MerR-family) was chosen for the following experiments. The fact that it is the only predicted transcription activator from S1C5 that was submitted for transcritption analysis together with its apparent basal transcription, similarly to the structural gene *SNA01225,* justifies its selection.

Lastly, concerning S5C20, *SNA06246* (LuxR-family) was the selected gene to be overexpressed. SNA06246 was only predicted positive transcription regulator from S5C20 submitted for transcritption analysis that presented low transcription levels, as the structural gene *SNA06257*.

3.2.1 Strategy

To overexpress selected CSR, pIB139 plasmid was preferred as the expression vector (Wilkinson *et al.*, 2002). pIB139 is an integrative vector that will allow cloning the selected regulatory genes under the control of the strong heterologous constitutive *ermE**p of *Saccharopolyspora erythraea*. Resulting plasmids were integrated into the *attB* site of the chromosome of *S. natalensis* to create strains overexpressing the CSRs. In order to clone CSRs sequences into pIB139, they were first amplified by PCR, using chromosomal DNA from *Streptomyces natalensis* ATCC 27448 as template. The used primers (Table 1, Materials & methods) contained *NdeI* or *EcoRV* restriction sites. The resulting PCR products were cloned into the pGEM®-T Easy vector and sequenced at STAB VIDA. Validated strains were then used for plasmid recovery and enzymatic digestion, with *NdeI* and *EcoRV*. Finally, CSR genes were purified and cloned in pIB139.

The obtained constructs, named pIBSNA01173, pIBSNA01239 and pIBSNA06246, were introduced into *E. coli* ET12567 [pUZ8002] before being transferred to *S. natalensis* ATCC 27448 (wild-type) and *S. natalensis* Δ*pimM* by intergeneric conjugation. The strains *S.*

natalensis [pIBSNA01173], *S. natalensis* [pIBSNA01239], *S. natalensis* [pIBSNA06246], SNA Δ*pimM* [pIBSNA01173], *S. natalensis* Δ*pimM* [pIBSNA01239] and *S. natalensis* Δ*pimM* [pIBSNA06246] were generated.

The empty pIB139 vector was also inserted into *E. coli* ET12567 [pUZ8002] and transferred then to *S. natalensis* Δ*pimM* strain. The resulting strain was used as control together with *S. natalensis* [pIB139] available at the laboratory.

Figure 27 summarizes the applied strategy.

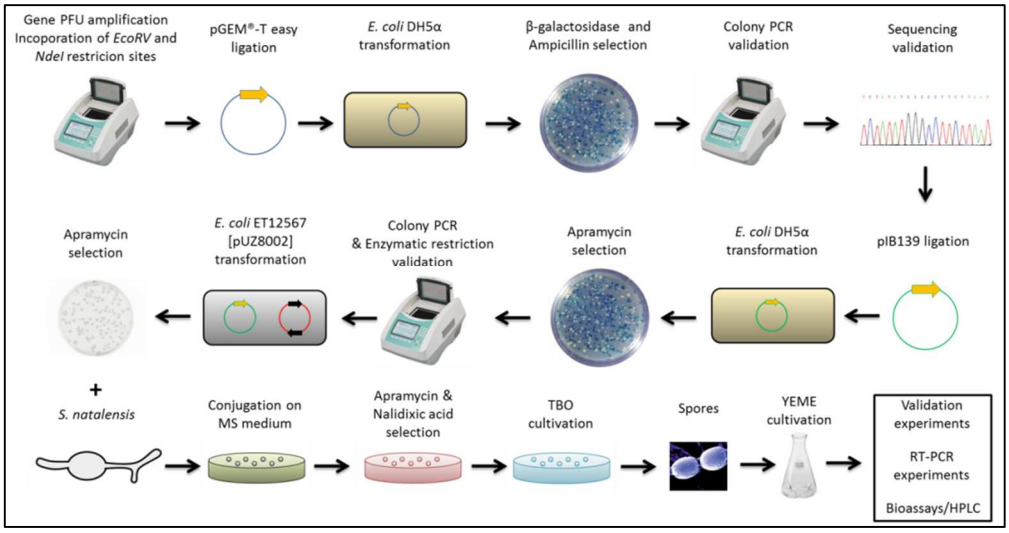

Figure 27 – Schematic representation of the chosen approach for the overexpression of the selected genes in *S. natalensis* and *S. natalensis* Δ*pimM.*

3.2.2 Exconjugants validation

To confirm the presence of the genes of interest three different clones for each conjugation were submitted for PCR analysis, using genomic DNA as template. The different sets of primers consisted in one that hybridises within the plasmid and another that hybridises in the gene pretended to be overexpressed. This way, it was possible to ascertain the presence of the construct within the genome of the strain.

Results of the PCR analysis are displayed in Figure 28. All the presented strains were validated since their amplicons matched their expected size (Table 3).

Table 3 – Expected sizes of the amplicons of the different *S. natalensis* (SNA) strains.

Strain	Primer pair	Expected size of the amplicon (bp)
SNA Δ*pimM* [pIB139]	pIB_U_F and pIB_U_R	293
SNA [pIBSNA01173] and SNA Δ*pimM* [pIBSNA01173]	pIB_U_F and S1C4R1_AS	400
SNA [pIBSNA01239] and SNA Δ*pimM* [pIBSNA01239]	pIB_U_F and S1C5R2_AS	491
SNA [pIBSNA06246] and SNA Δ*pimM* [pIBSNA06246]	pIB_U_F and S5C20R4_AS	541

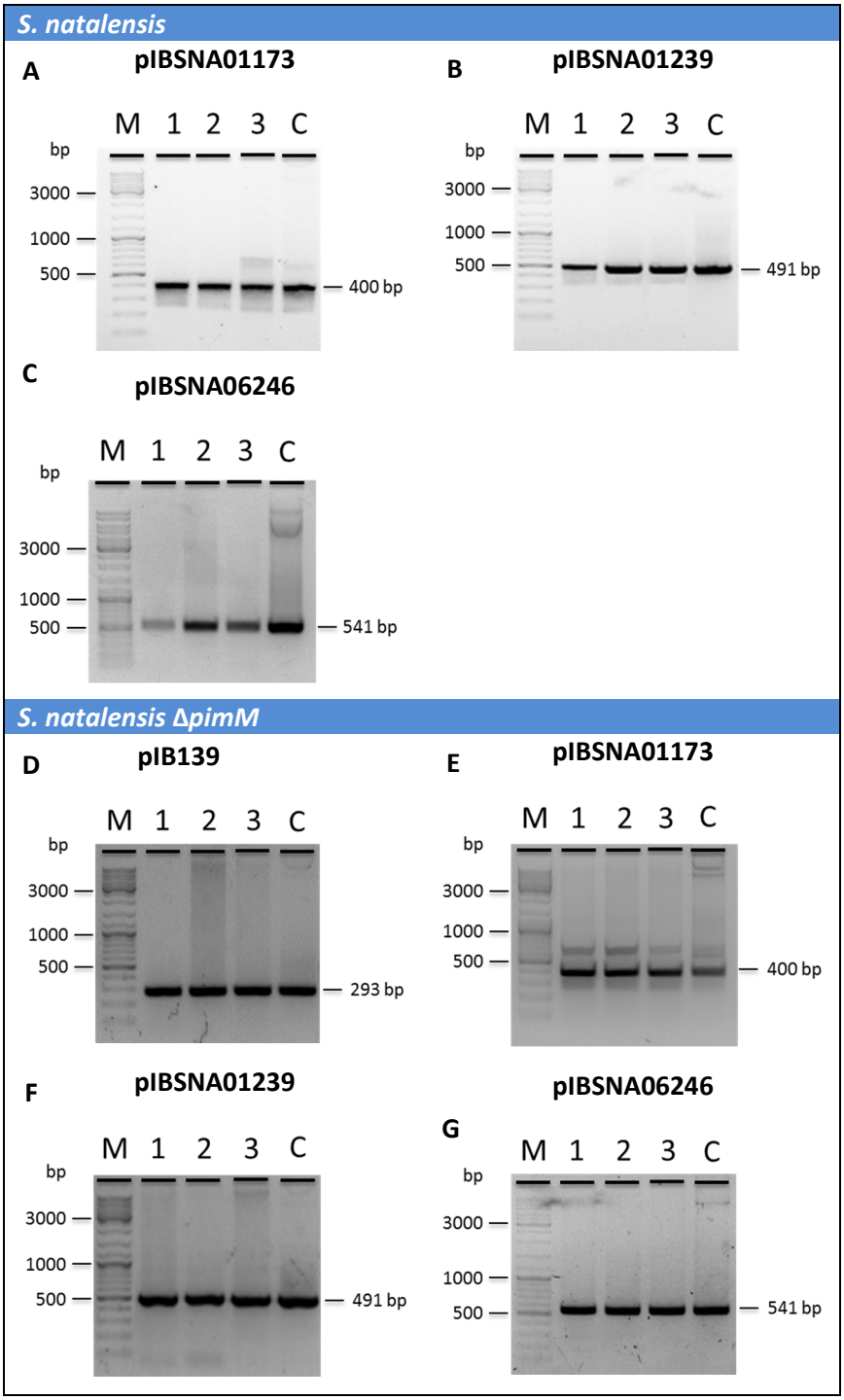

Figure 28 – PCR validation of exconjugants. Three exconjugants were randomly selected (lanes 1, 2 and 3) for each conjugation: A and E – pIBSNA01173; B and F – pIBSNA01239; C and G – pIBSNA06246; D – pIB139. Plasmid DNA was used as control (lane C) and Gene Ruler (Thermo Scientific; #SM0331) was used as molecular-weight size marker (M).

Additionally, three clones for each conjugation already validated by PCR were also validated by for Southern blotting analysis (Fig. 29-32). Internal fragments of the studied genes (*SNA01173*, *SNA01239* and *SNA06246*) or of the pIB139 plasmid were used as probes. Restriction enzymes were selected in a way to obtain a digested fragment that should include part of the surrounding region of pIB139 (and derivative constructs) integration sites at the

chromosome (*attB*) together with the probes hybridization sites. Since unspecific bands of plasmid pSNA1 can be detected, its restriction fragments were also taken into consideration (Table 4).

Table 4 – Expected sizes for southern hybridization bands of the different *S. natalensis* (SNA) strains. (WT, wild-type; NA, not applicable).

Target	Used restriction enzyme(s)	Strain	Expected size of the gDNA bands (bp)	Expected size of pSNA1 bands (bp)
pIB139	*Bgl*II	SNA WT and SNA Δ*pimM*	NA	5969; 3398
		SNA [pIB139] and SNA Δ*pimM* [pIB139]	11517	
SNA01173	*Eco*RV and *Kpn*I	SNA WT and SNA Δ*pimM*	8671	3677; 3464; 1222; 1004
		SNA [pIBSNA01173] and SNA Δ*pimM* [pIBSNA01173]	7585	
SNA01239	*Age*I and *Not*I	SNA WT and SNA Δ*pimM*	4409	3074; 1584; 1563; 885; 827; 781; 437; 153; 45; 15
		SNA [pIBSNA01239] and SNA Δ*pimM* [pIBSNA01239]	7813	
SNA06246	*Nco*I	SNA WT and SNA Δ*pimM*	7962	9128; 239
		SNA [pIBSNA06246] and SNA Δ*pimM* [pIBSNA06246]	6475	

S. natalensis [pIB139], *S. natalensis* [pIBSNA01173] exconjugant #8, *S. natalensis* [pIBSNA01239] exconjugant #3, *S. natalensis* [pIBSNA06246] exconjugant #7, *S. natalensis* Δ*pimM* [pIB139] exconjugant #4, *S. natalensis* Δ*pimM* [pIBSNA01173] exconjugant #11, *S. natalensis* Δ*pimM* [pIBSNA01239] exconjugant #3, *S. natalensis* Δ*pimM* [pIBSNA06246] exconjugant #1 were selected for the remaining experiments. These strains are named, from here on, without the cardinal and associated number, as representatives of each conjugation.

3.2.3 Transcription analysis of the exconjugants

Due to time constraints, transcription analyses were performed using *S. natalensis* wild-type and derived exconjugation strains. Two biological replicates of each strain were used in order to ensure the liability of the results, with the exception for *S. natalensis* [pIB139] due to technical issues.

Obtained results are similar between S1C5 and S5C20 biosynthetic gene clusters: the CSR chosen to be overexpressed showed increased levels when compared with the control strains (Fig. 33 and 34). However, the studied structural gene did not presented any increased level between the exconjugants and control strains.

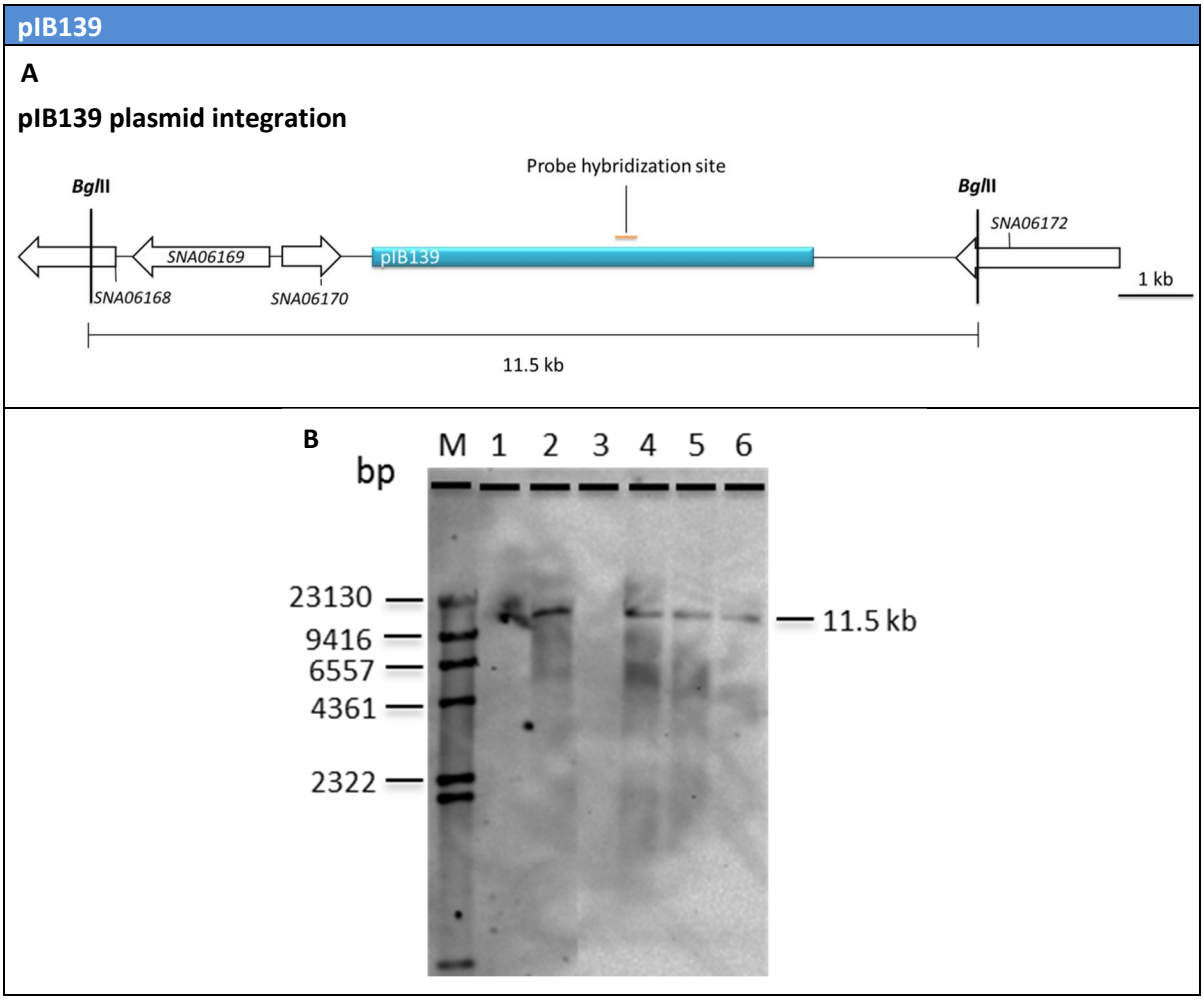

Figure 29 – Southern blot confirmation of strains *S. natalensis* [pIB139] and *S. natelensis* Δ*pimM* [pIB139] genetic identity. A). *Bgl*II restriction pattern for the predicted pIB139 integration (shown in a blue box). The probe used for Southern hybridisation is indicated by the orange line. B). Southern blot hybridization of digested *Streptomyces* exconjugants genomic DNA by making use of dioxygenin labelled probes. λ HindIII (Thermo Scientific; #SM0102) was used as molecular-weight size marker (M). Lane 1 – *S. natalensis*; 2 - *S. natalensis* [pIB139]; 3 - *S. natalensis* Δ*pimM*; 4 – *S. natalensis* Δ*pimM* [pIB139] exconjugant #3; 5 – *S. natalensis* Δ*pimM* [pIB139] exconjugant #4; 6 – *S. natalensis* Δ*pimM* [pIB139] exconjugant #5.

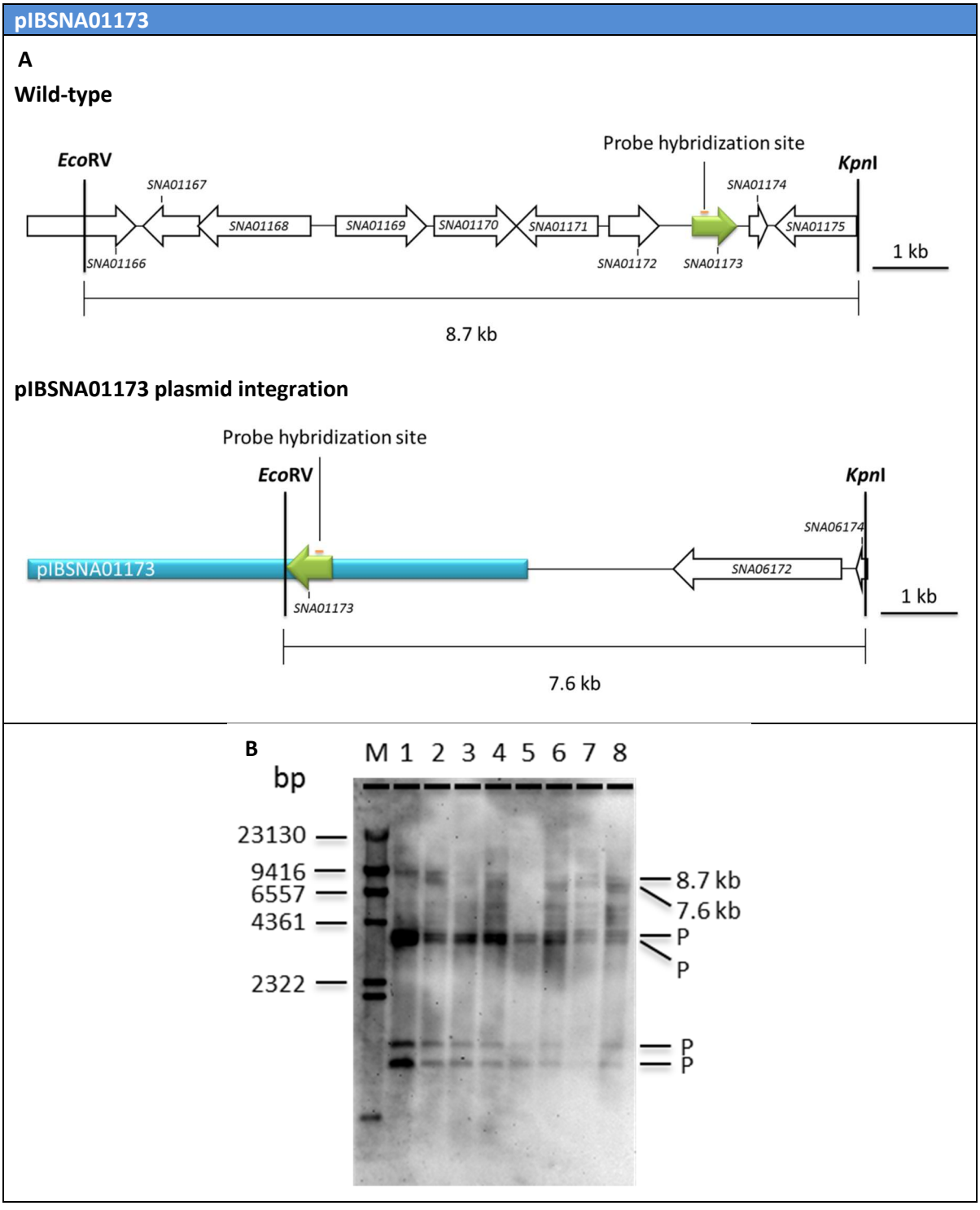

Figure 30 – Southern blot confirmation of strains *S. natalensis* [pIBSNA01173] and *S. natelensis* Δ*pimM* [pIBSNA01173] genetic identity. A). *Eco*RV and *Kpn*I restriction pattern for the predicted pIBSNA01173 integration (shown in a blue box). The CSR gene *SNA01173* is represented by a green arrow. The probe used for Southern hybridisation is indicated by the orange line. B). Southern blot hybridization of digested *Streptomyces* exconjugants genomic DNA by making use of dioxygenin labelled probes. λ HindIII (Thermo Scientific; #SM0102) was used as molecular-weight size marker (M). Lane 1 – *S. natalensis*; 2 - *S. natalensis* [pIBSNA01173] exconjugant #8; 3 - *S. natalensis* [pIBSNA01173] exconjugant #9; 2 - *S. natalensis* [pIBSNA01173] exconjugant #10; 5 - *S. natalensis* Δ*pimM*; 6 – *S. natalensis* Δ*pimM* [pIBSNA01173] exconjugant #5; 7 – *S. natalensis* Δ*pimM* [pIBSNA01173] #6; 8 – *S. natalensis* Δ*pimM* [pIBSNA01173] exconjugant #11. P – Unspecific band from the plasmid pSNA1.

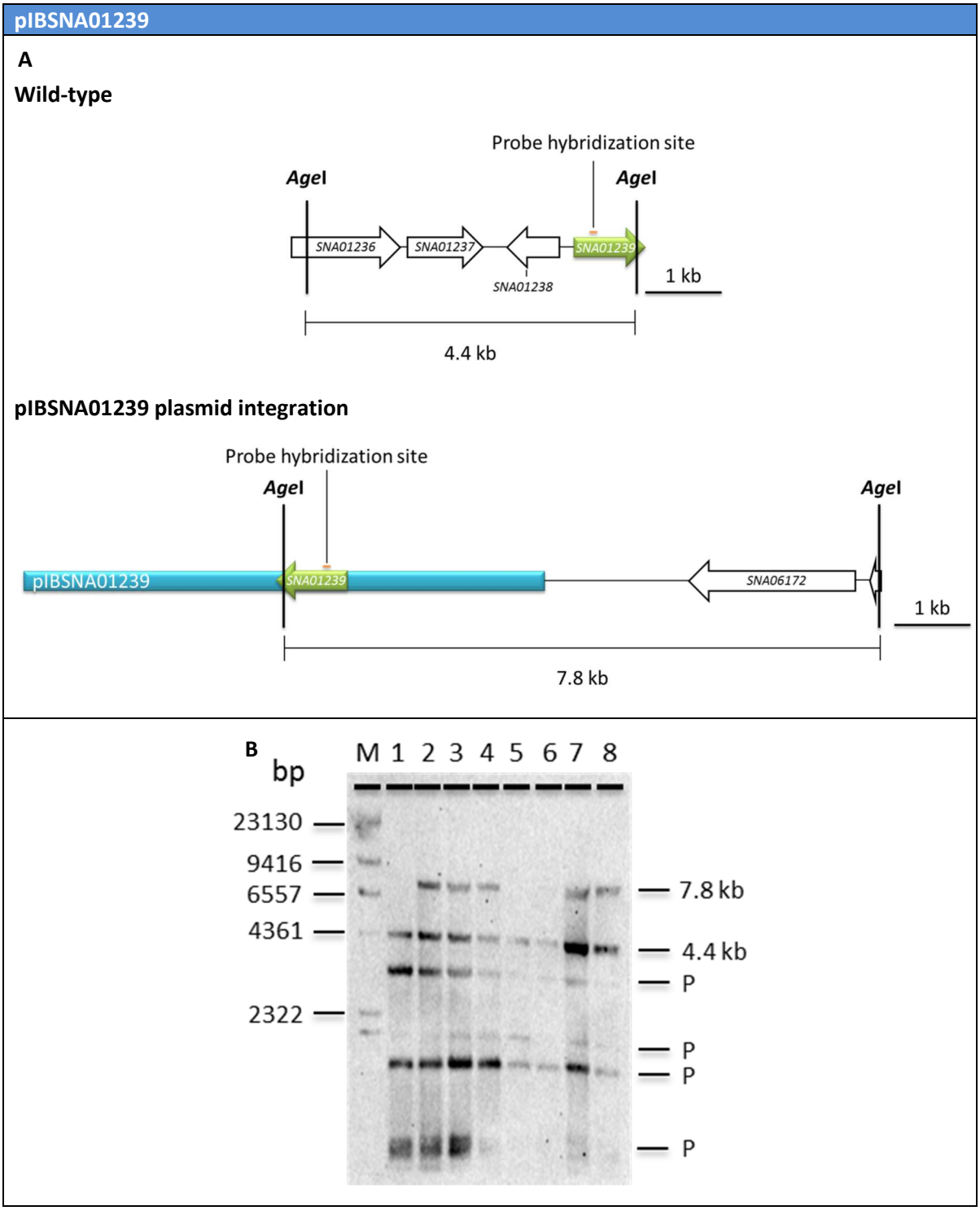

Figure 31 – Southern blott confirmation of strains *S. natalensis* [pIBSNA01239] and *S. natelensis* Δ*pimM* [pIBSNA01239] genetic identity. A). *Age*I and *Not*I restriction pattern for the predicted pIBSNA01239 integration (shown in a blue box). The CSR gene *SNA01239* is represented by a green arrow. The probe used for Southern hybridisation is indicated by the orange line. B). Southern blot hybridization of digested *Streptomyces* exconjugants genomic DNA by making use of dioxygenin labelled probes. λ HindIII (Thermo Scientific; #SM0102) was used as molecular-weight size marker (M). Lane 1 – *S. natalensis*; 2 - *S. natalensis* [pIBSNA01239] exconjugant #2; 3 - *S. natalensis* [pIBSNA01239] exconjugant #3; 2 - *S. natalensis* [pIBSNA01239] exconjugant #9; 5 - *S. natalensis* Δ*pimM*; 6 – *S. natalensis* Δ*pimM* [pIBSNA01239] exconjugant #6; 7 – *S. natalensis* Δ*pimM* [pIBSNA01239] exconjugant #3; 8 – *S. natalensis* Δ*pimM* [pIBSNA01239] exconjugant #5. P – Unspecific band from the plasmid pSNA1.

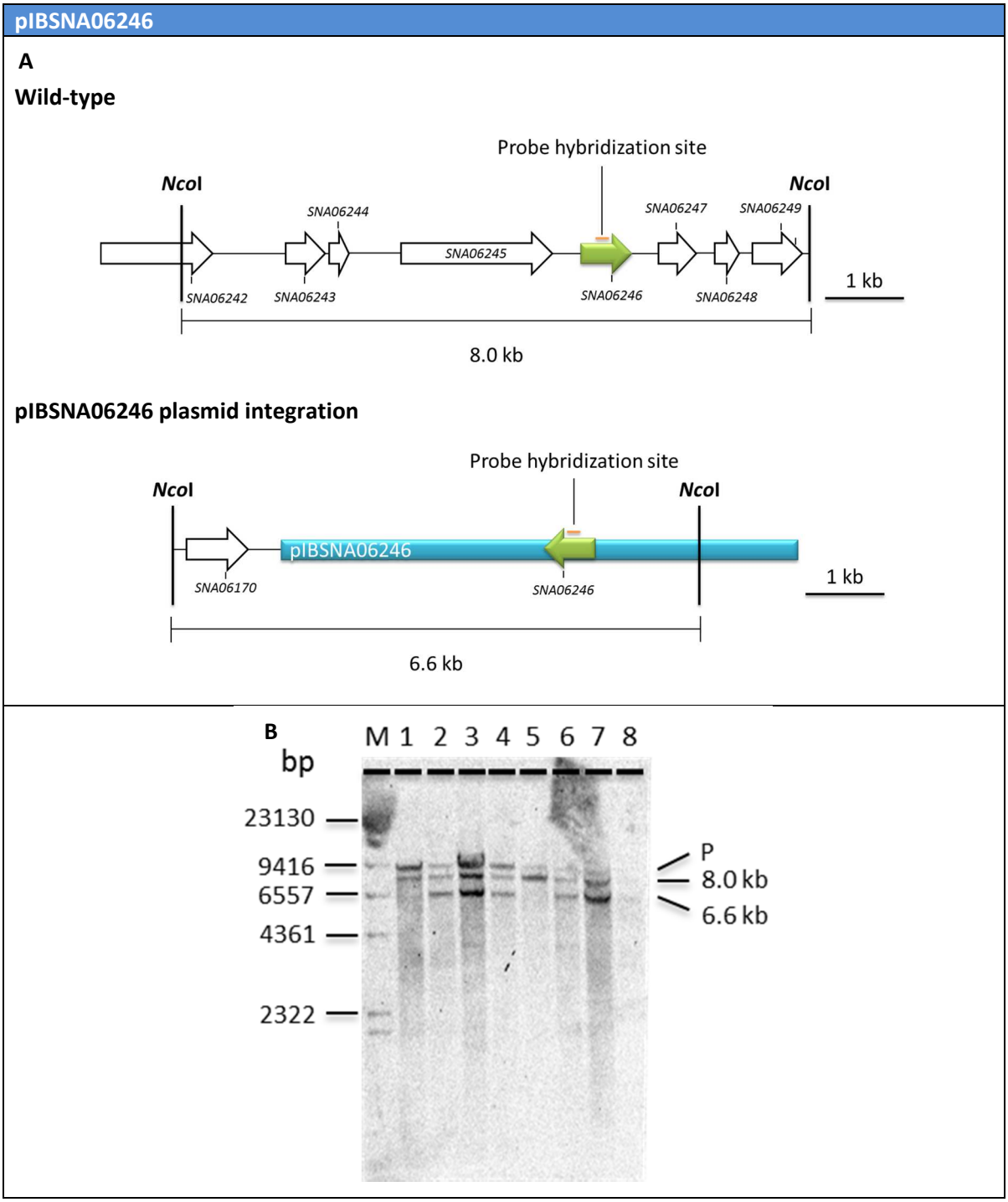

Figure 32 – Southern blott confirmation of strains *S. natalensis* [pIBSNA06246] and *S. natelensis* Δ*pimM* [pIBSNA06246] genetic identity. A). *Age*I and *Not*I restriction pattern for the predicted pIBSNA01239 integration (shown in a blue box). The CSR gene *SNA06246* is represented by a green arrow. The probe used for Southern hybridisation is indicated by the orange line. B). Southern blot hybridization of digested *Streptomyces* exconjugants genomic DNA by making use of dioxygenin labelled probes. λ HindIII (Thermo Scientific; #SM0102) was used as **m**olecular-weight size marker (M). Lane 1 – *S. natalensis*; 2 - *S. natalensis* [pIBSNA06246] exconjugant #5; 3 - *S. natalensis* [pIBSNA06246] exconjugant #7; 2 - *S. natalensis* [pIBSNA006246] exconjugant #17; 5 - *S. natalensis* Δ*pimM*; 6 – *S. natalensis* Δ*pimM* [pIBSNA06246] exconjugant #1; 7 – *S. natalensis* Δ*pimM* [pIBSNA06246] exconjugant #5; 8 – *S. natalensis* Δ*pimM* [pIBSNA06246] exconjugant #8. P – Unspecific band from the plasmid pSNA1.

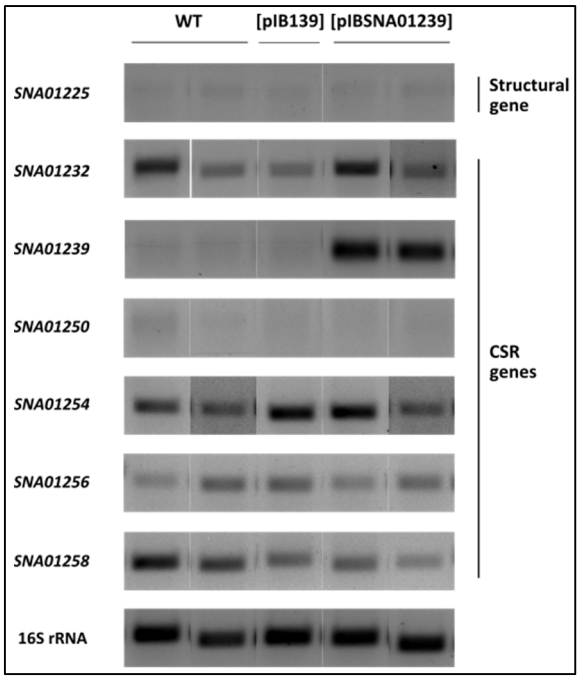

Figure 33 – Transcription analysis of S1C5 selected genes by RT-PCR. Transcripts of *SNA01225* (NRPS encoding gene), *SNA01232* (TetR-family encoding gene), *SNA01239* (MerR-family encoding gene), *SNA01250* (DeoR-family encoding gene), *SNA01254* (PaaX-family encoding gene), *SNA01256* (TetR-family encoding gene), *SNA01258* (TetR-family encoding gene), 16S rRNA (positive control) and *pimS1* (positive control for WT and negative control for Δ*pimM*) were analysed by RT-PCR with 30 cycles on cDNA generated from RNA isolated at early stationary phase. Two biological replicates are shown for WT and [pIBSNA01239].

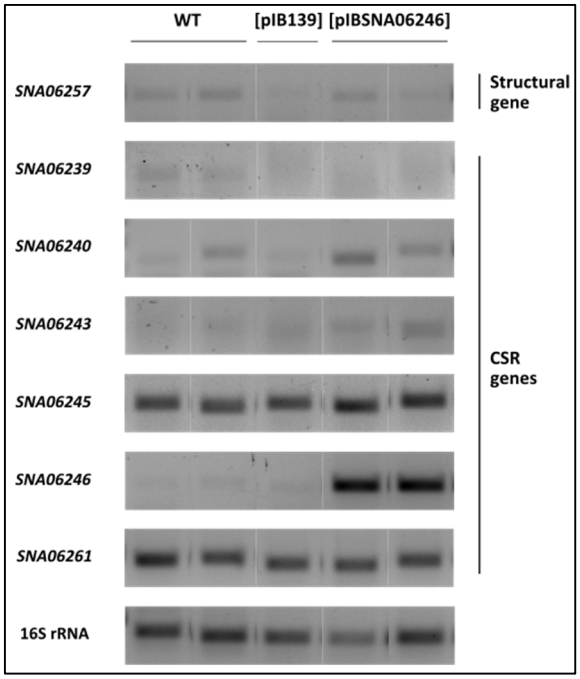

Figure 34 – Transcripton analysis of S5C20 selected genes by RT-PCR. Transcripts of *SNA06257* (mixed PKS-NRPS encoding gene), *SNA06239* (GntR-family encoding gene), *SNA06240* (LysR-family encoding gene), *SNA06243* (AsnC-family encoding gene), *SNA06245* (AfsR-family encoding gene), *SNA06246* (LuxR-family encoding gene), *SNA06261* (serine-threonine kinase encoding gene), 16S rRNA (positive control) and *pimS1* (positive control for WT and negative control for Δ*pimM*) were analysed by RT-PCR with 30 cycles on cDNA generated from RNA isolated at early stationary phase. Two biological replicates are shown for WT and [pIBSNA06246].

The gene expression analysis indicates that the CSR overexpression was successful although apparently the expression of structural genes was not activated. The results suggest that these clusters might have additional layers of regulation beyond the selected genes.

Worth to note that secondary metabolism gene clusters often possess more than a single associated CSR, wherein one of the CSR is an ultimate regulator of antibiotic production, while others may act either singularly on the ultimate regulatory gene, or pleiotropically on unrelated and unlinked genes (Makitrynskyy *et al.*, 2013). Thus, it is possible that there are regulatory cascades controlling the activation of the biosynthetic gene clusters. By way of example, in *S. coelicolor*, the cryptic polyketide (CPK) biosynthesis depends upon the *cpkO* activation. Conversely, the activation of the *cpk* cluster also involves, among other regulatory molecules, the protein ScbR, coded by *scrB*, and gamma-butyrolactone congeners (SCB) coded by *scbA*. On the one hand, ScbR binds to *cpkO* promoter region, preventing its transcription. On the other hand, SCB is able to bind to ScbR, sequestering all ScbR free proteins and releasing *cpkO* from repression (Liu *et al.*, 2013).

This way, it is conceivable that SNA01239 and SNA06246 are only part of a regulatory cascade in which other negative CSR, or even some other pleiotropic regulators located outside the cluster, are preventing them from activating the structural genes.

As mentioned in the introduction there are nutrient-sensing regulators already ascribed to have a role in the antibiotic production in *S. coelicolor.* Indeed, phenomena like carbon catabolite repression may have counteracted the effect of the overexpression of these CSRs. Moreover, the media may lack an important co-factor needed for the transcription regulators to acquire the DNA binding conformation (Thomas and Chiang, 2006).

Lastly, *SNA01239* and *SNA06246* may not be positive regulators, but be negative instead. Either the MerR-family (Brown *et al.*, 2003) or the LuxR-family (Chen and Xie, 2011) have been reported to have negative transcription regulators as members.

In the case of S1C4, the putatively overexpressed gene (*SNA01173*) shows apparent similar expression levels as the controls (Fig. 35). Such results are inconsistent with the ones obtained for the first transcription analysis, in which *SNA01173* showed low transcription levels (Fig. 18). That could be due to the fact that *SNA01173* holds a different transcription pattern along *S. natalensis* growth. Nieselt and co-workers have already shown that genes composing different biosynthetic clusters have their own transcription patterns, in which the expression varies along time (Nieselt *et al.*, 2010). Thus, in spite of all mycelium samples were harvested at the beginning of the stationary phase, it may have happened that they were in slightly different growth phase stages. Moreover, as already mentioned in this work, bacteriocins are all produced during the primary phase of growth, though antibiotics are

usually secondary metabolites (Beasley and Saris, 2004). This fact emphasises the difference that can be seen in gene transcription patterns when samples are collected at different stages of growth. Furthermore, the overexpression of the gene, which is not evident in the RT-PCR results, did not allow the activation of the studied structural gene. Similarly to the other clusters it could be (1) due to the effect of CSR or other pleiotropic regulators acting as repressors, (2) to the lack of a cofactor, (3) to the used media or (4) SNA01173 acts as a negative regulator instead. Regarding the latter point, it should be noted that there are secondary metabolite biosynthetic gene clusters with no associated CSR, namely moenomycin (Ostash *et al.*, 2007).

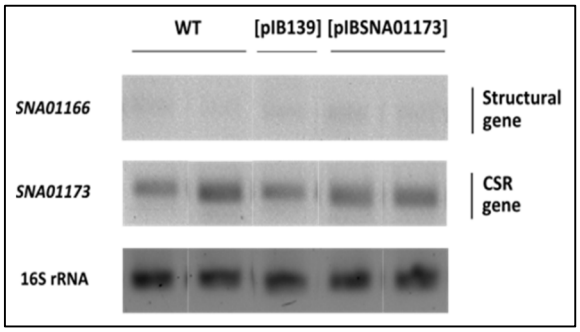

Figure 35 – Transcription analysis of S1C4 selected genes by RT-PCR. Transcripts of *SNA01166* (TAP encoding gene), *SNA01173* (LuxR-family encoding gene) and 16S rRNA (positive control) were analysed by RT-PCR with 30 cycles on cDNA generated from RNA isolated at early stationary phase. Two biological replicates are shown for WT and [pIBSNA01173].

In addition to the mentioned four reasons, it should be mentioned that, unlike S1C5 and S5C20 studied structural genes, it might be possible that SNA01166 is not involved in the biosynthesis of the end bacteriocin. One other protease, potentially coded by a predicted hypothetical protein, might be the one responsible for the cleavage of the N terminal of the pre-peptide.

3.2.4 Comparative metabolomics

3.2.4.1 Bioassay experiments

Despite the fact that the activation of the expression of the studied genes was apparently not achieved, all new developed strains were submitted for a preliminary comparative metabolic profile. 250 µL of culture broth was concentrated and tested against a panel of test organisms that includes reference strains of Gram-positive bacteria (*Bacillus cereus, Staphylococcus aureaus* and *Enterococcus faecalis*), Gram-negative bacteria (*Escherichia coli*) and yeast (*Saccharomyces cerevisiae*).

No inhibition halos were detected in the bioassays performed with any of the tested Gram-positive bacteria and Gram-positive negative bacteria. This could mean that none of the *S. natalensis* strains produce any metabolite capable of killing these microorganisms or

suppressing their growth. Otherwise, the used amount of media supernatant does not contain the minimum concentration of the putative product so it can work as an inhibitory agent.

Regarding bioassays performed with *Saccharomyces cerevisiae* as test organism it is possible to notice the presence of inhibition halos surround the area with media supernatant from *S. natalensis* wild-type and derivative strains (Fig. 36). *S. natalensis* produces pimaricin, na antifungal that inhibits the growth of the yeast *S. cerevisiae*, hence the inhibition halos. The measurement of inhibition halos revealed differences between the applied supernatants. pIB139 vector incorporation appears to slightly decrease the production of the metabolite(s) responsible for the inhibition zone. Moreover, while no differences were detected between the halos from *S. natalensis* [pIB139] and *S. natalensis* [pIBSNA06246], *S. natalensis* [pIBSNA01239] culture broth led to lower inhibition halos. These results suggest different production titers of pimaricin.

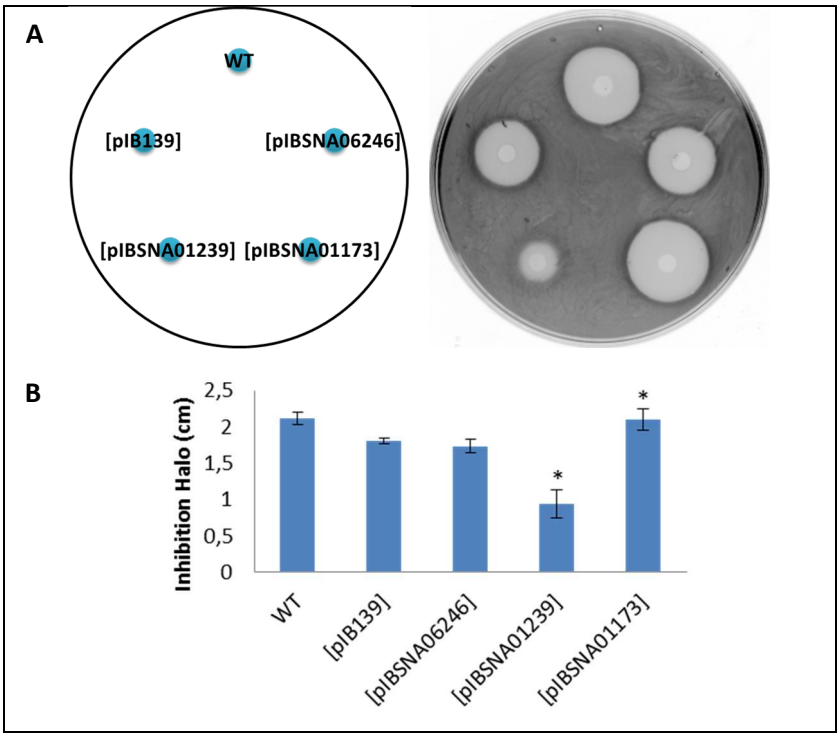

Figure 36 – Bioassay results for the collected culture medium, using *Saccharomyces cerevisiae* as test organism. A). Plate showing antibacterial activity of *S. natalensis* and developed exconjugants broth cultures by agar well diffusion method. B). Graphical representation of the inhibition halos. Results show the average obtained among two technical replicates of two independent biological replicates. Vertical bars indicate standard deviation of the mean values. The differences between exconjugants *S. natalensis* [pIBSNA01239] or *S. natalensis* [pIBSNA01239] and *S. natalensis* [pIB139] samples were assessed by independent t-test. (*, $P < 0.05$).

Contrasting with the latter results, no inhibition zones were visualized in the bioassays performed with supernatants collected from culture broth of *S. natalensis* Δ*pimM* and derivative strains that do not produce pimaricin (data not shown). Altogether, bioassay results with *Saccharomyces cerevisiae* as test organism suggest that pimaricin is the compound that is

supressing the yeast growth and no other inhibitory agent is being produced as the result of overexpressing *SNA06246*, *SNA01239* and *SNA01173*. Nevertheless, these results do not represent quantitative data.

3.2.4.2 HPLC experiments

In order to clarify the previous results and look for possible new products that can be produced by the developed exconjugants to which the tested microorganisms are not sensible, culture broth methanolic extracts were analysed by HPLC.

The resulting diode array spectra for all strains are displayed in Appendix II. Analysis of the wild-type extract at 304 nm revealed a main peak with a retention time (RT) of 13.85 min (peak #1, Fig. 37). This peak had the same RT as the pimaricin standard and displayed the typical absorption spectrum of pimaricin with three characteristic absorption maximums (290, 304 and 319 nm). A second peak, with a RT of 16.6 min also displayed a tetraene typical absorption spectrum (peak #2, Fig. 37). This peak displayed an additional absorption peak at 260 nm that suggests the presence of a conjugated double bond at C2-C3 and C4-C5 in the pimaricin backbone (Mendes *et al.*, 2005). The introduction of the epoxy group by the P450 monooxygenase into the pimaricin module at C4-C5 is the last step of pimaricin biosynthesis (Mendes *et al.*, 2005) (Fig. 38). Therefore, the presence of peaks displaying an additional absorption maximum at 260 nm suggests the presence of pimaricin precursor in the wild-type culture broth.

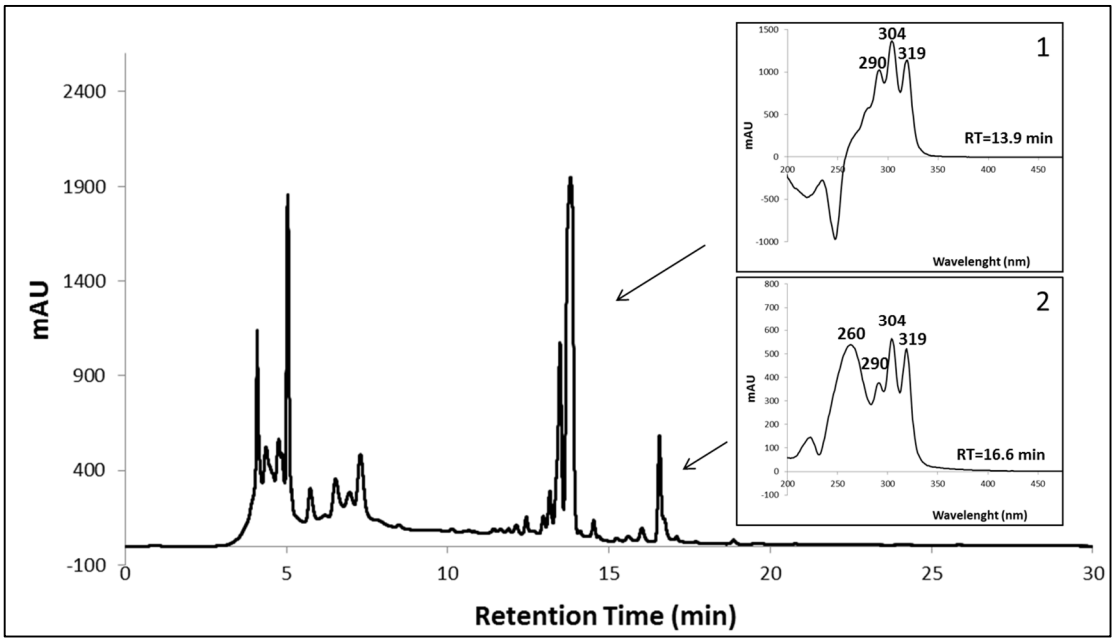

Figure 37 – HPLC chromatogram at 304 nm of *S. natalensis* culture broth extracts. 1). Pimaricin peak spectra. 2). Putative pimaricin precursor. Arrows indicate peaks corresponding absorption spectra.

Nevertheless, the main differences were detected at 400 nm between the different wild-type derivative exconjugants (Fig. 39-41).

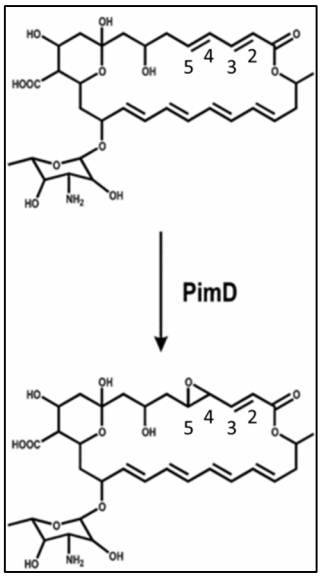

Figure 38 – Last step for pimaricin biosynthesis: de-epoxypimaricin conversation into pimaricin, catalysed by PimD [Adapted from Mendes *et al*., 2005].

The first difference refers to different intensities of two retention peaks among the analysed strains: peak #3 (RT of 14.2 min) and peak #5 (RT of 18.9 min). These peaks were both increased in strains *S. natalensis* [pIBSNA016246] and *S. natalensis* [pIBSNA01173] and are not present in *S. natalensis* [pIBSNA01239] when compared to the control strain (*S. natalensis* [pIB139]). The peak area was normalized against the culture dry weight in order to be able to perform a comparative analysis (Fig. 42). Peak #3 had an 11.3 and 5.3 fold increase in strains *S. natalensis* [pIBSNA01173] and *S. natalensis* [pIBSNA06246], respectively. The increase in peak #5 was not as pronounced as peak #3, although it led to 1.85 and 1.84 fold increase in strains *S. natalensis* [pIBSNA01173] and *S. natalensis* [pIBSNA06246], respectively.

The second main difference between the analysed strains was the presence of three new retention peaks at 400 nm: peak #4 (RT of 17.5 min), peak #6 (RT of 13.2 min) and peak #7 (RT of 13.3 min). Peaks #4 and #6 are present in the culture broth of strains *S. natalensis* [pIBSNA01239] and *S. natalensis* [pIBSNA01173]. Strain *S. natalensis* [pIBSNA06246] displayed small peak at the corresponding RT. Finally peak #7 was present exclusively in strain *S. natalensis* [pIBSNA01173].

Worth note that the absorption spectra of all peaks had similar absorption maximums, ranging from 402 to 406 nm. Additionally the absorption spectra of peak #6 and #7 displayed typical pimaricin spectra. This is due to the proximity of pimaricin or its precursors/by-products RT.

Concerning Δ*pimM* and derivative strains, no differences were detected between them (Fig. 43-45). Nonetheless, it is interesting to observe that none of the new obtained spectra displayed the pimaricin typical spectrum between 250 and 350 nm. These results reinforce the

assertion that of both #6 and #7 in wild-type derivative strains consist in a mixture of two independent spectra. As absorption spectra were similar peaks with same RT they were named equally.

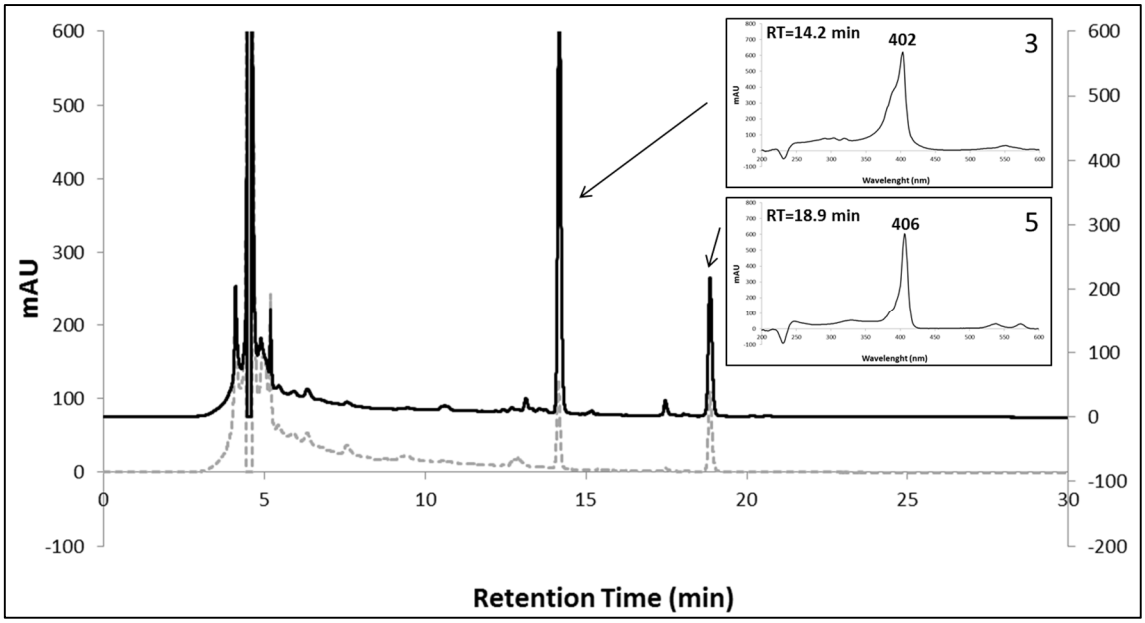

Figure 39 – HPLC chromatogram at 400 nm of *S. natalensis* [pIBSNA06246] (black line) and *S. natalensis* [pIB139] (dashed grey line) culture broth extracts. Arrows indicate peaks corresponding absorption spectra.

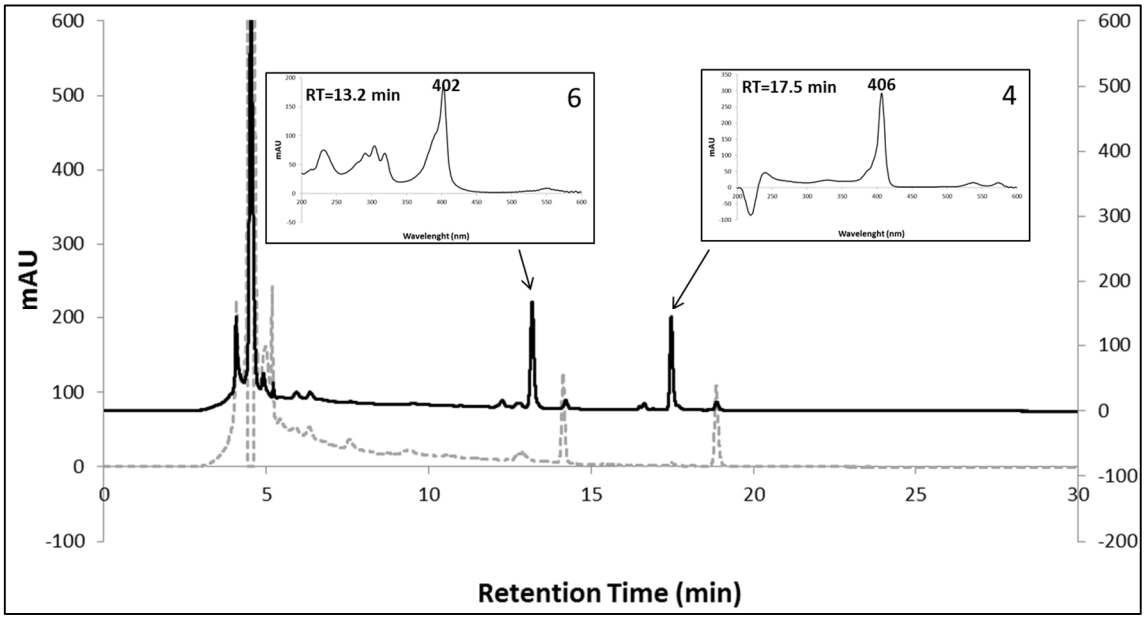

Figure 40 – HPLC chromatogram at 400 nm of *S. natalensis* [pIBSNA01239] (black line) and *S. natalensis* [pIB139] (dashed grey line) culture broth extracts. Arrows indicate peaks corresponding absorption spectra.

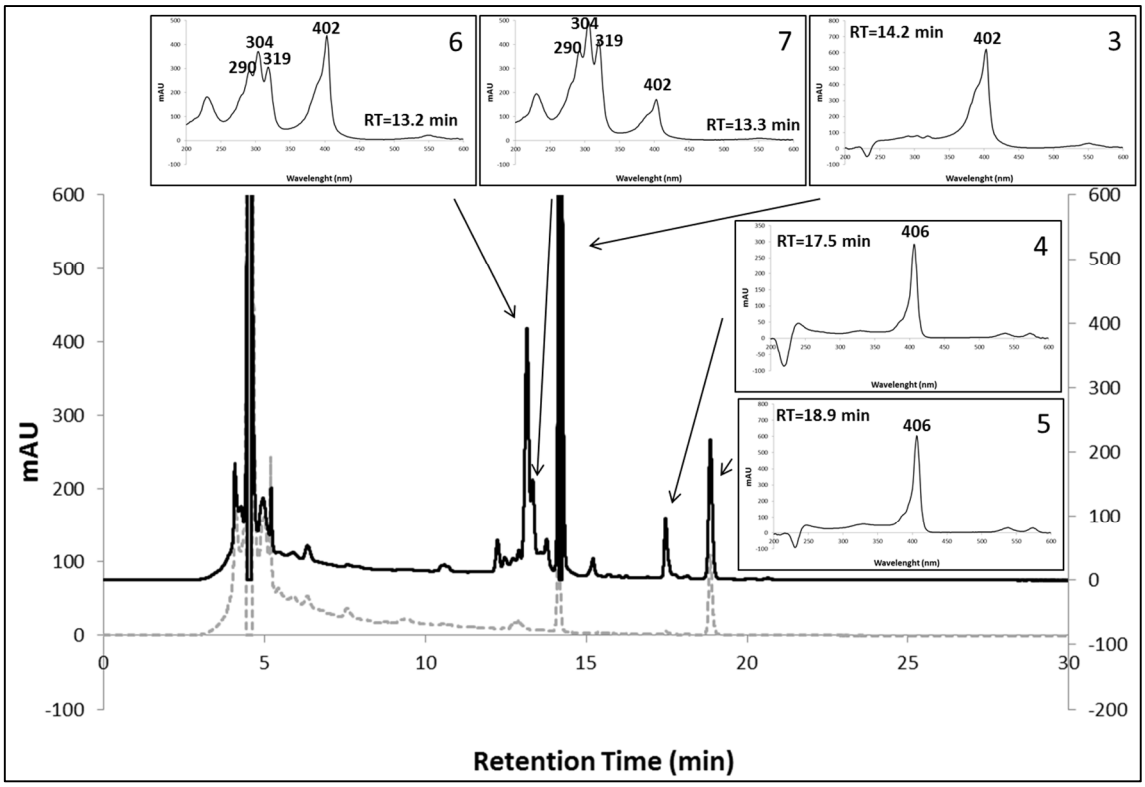

Figure 41 – HPLC chromatogram at 400 nm of *S. natalensis* [pIBSNA01173] (black line) and *S. natalensis* [pIB139] (dashed grey line) culture broth extracts. Arrows indicate peaks corresponding absorption spectra.

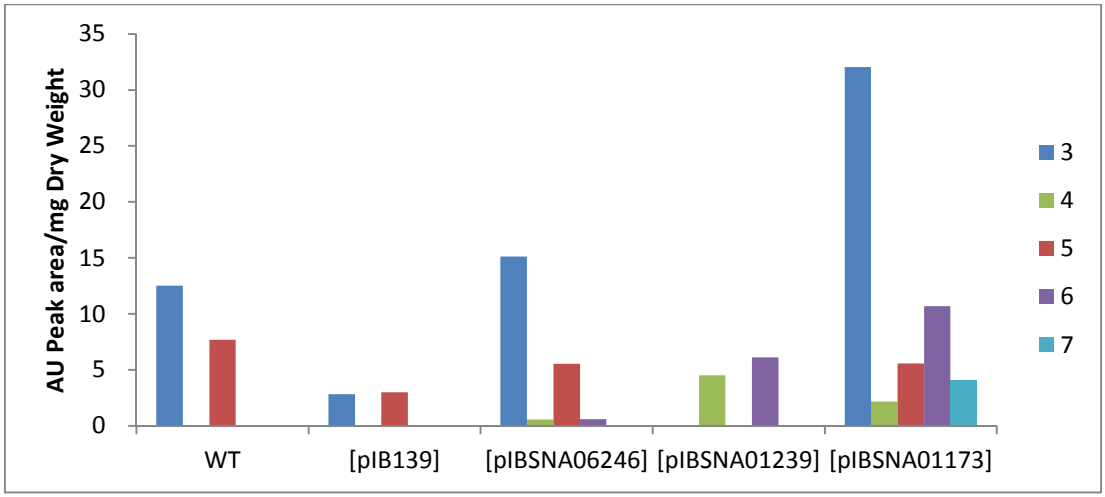

Figure 42 – Peak areas per dry weight (mg) ratios identified in HPLC chromatograms at 400 nm.

Moreover, it is also thought-provoking that peaks #4, #6 and #7 that are not present neither in the wild-type strain nor in *S. natalensis* [pIB139], are present in all Δ*pimM* strains chromatograms. This fact points that the silencing of a biosynthetic gene cluster does alter the metabolic profiling of a strain. Moreover, these results also support the idea that that the silencing of one gene cluster putatively sets free metabolic precursors essential for the production of other specialized metabolites.

Lastly, it should be mentioned that analysis at 304 nm also allowed to quantify differences in the pimaricin production between *S. natalensis* wild-type and derived strains

(Fig. 46). Results showed no differences between the wild-type strain and *S. natalensis* [pIB139], pointing that pIB139 vector integration did not affect the pimaricin production. Regarding the remaining exconjugants it is possible to ascertain that *S. natalensis* [pIBSNA06246] and *S. natalensis* [pIBSNA01239] showed decreased volumetric and specific pimaricin production values when compared with *S. natalensis* [pIB139]. In fact, *S. natalensis* [pIBSNA01239] presented significant decreases of 3.28 and 2.34 fold in volumetric and specific pimaricin production, respectively, when compared with *S. natalensis* [pIB139]. In turn, *S. natalensis* [pIBSNA01173] presented increased volumetric and specific pimaricin production values of 1.88 and 1.70 fold, respectively. The latter results, together with the ones obtained for *S. natalensis* [pIBSNA01173] bioassays (in which this strain exhibited larger inhibition halos than the control one) suggested that the overexpressed LuxR protein it is exerting a pleiotropic effect, modulating the pimaricin biosynthetic gene cluster. However, the results obtained for *S. natalensis* Δ*pim*M derivative strains (no pimaricin peak) showed that SNA01173 overexpression is not sufficient to bypass the absence of the pimaricin CSR PimM.

Summarising, the results obtained among all the strains point that they have different metabolic patterns. Two new peaks (#4 and #6) were detected in all *S. natalensis* strains overexpressing the different CSRs. Additionally *S. natalensis* [pIBSNA01173] appears to be overproducing pimaricin and the product corresponding to peak #7. In turn, *S. natalensis* [pIBSNA01239] underproduces pimaricin and does not even produce the compound related to peaks #3 and #5 present in the control strain. This way, *S. natalensis* [pIBSNA01239] could be deviating metabolic precursors for the production of a novel compound. *S. natalensis* [pIBSNA06246] was the new exconjugate with the most similar metabolic profile as *S. natalensis* [pIB139], the control strain. Lastly, no differences were detected between Δ*pimM* and derivative strains, although all have appeared to be producing the products that were not detected in the wild-type, *S. natalensis* [pIB139] or *S. natalensis* [pIBSNA01239]. Worth to note that all the similar absorption spectra suggest that the novel peaks could be by-products of one compound.

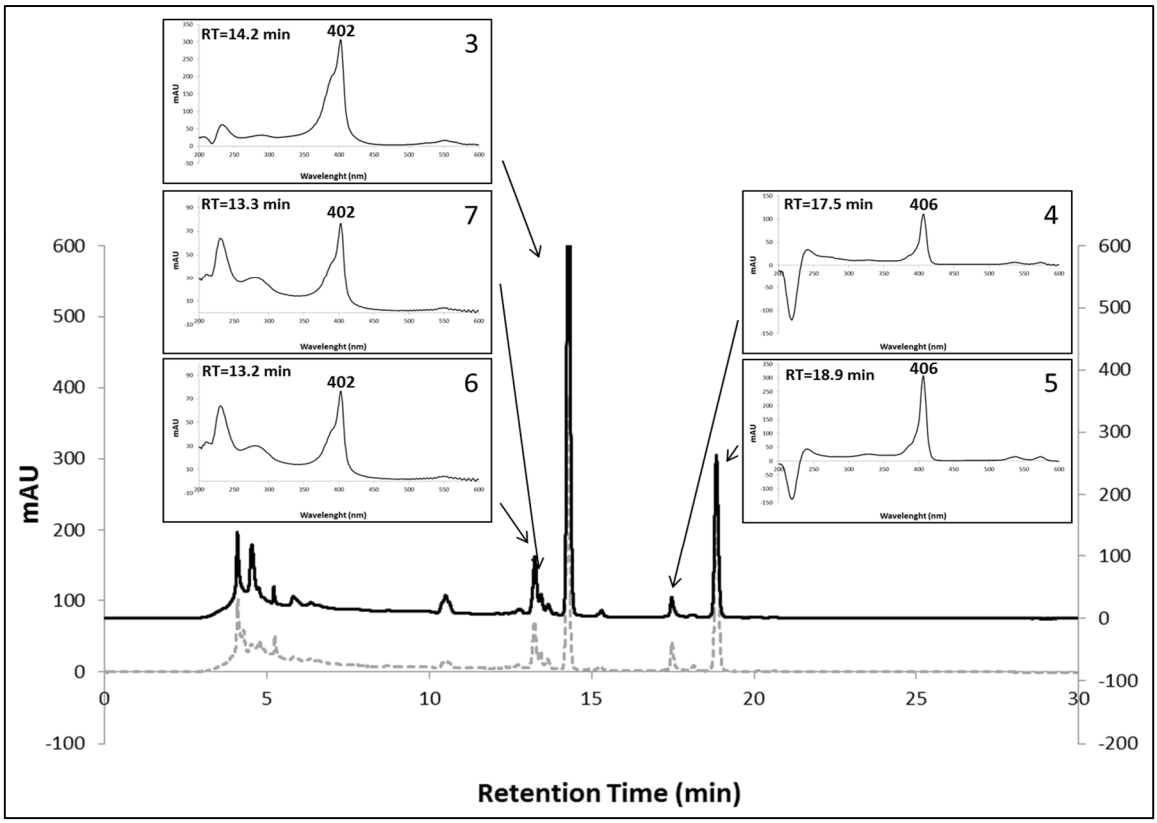

Figure 43 – HPLC chromatogram at 400 nm of *S. natalensis* Δ*pimM* [pIBSNA06246] (black line) and *S. natalensis* [pIB139] (dashed grey line) culture broth extracts. Arrows indicate peaks corresponding absorption spectra.

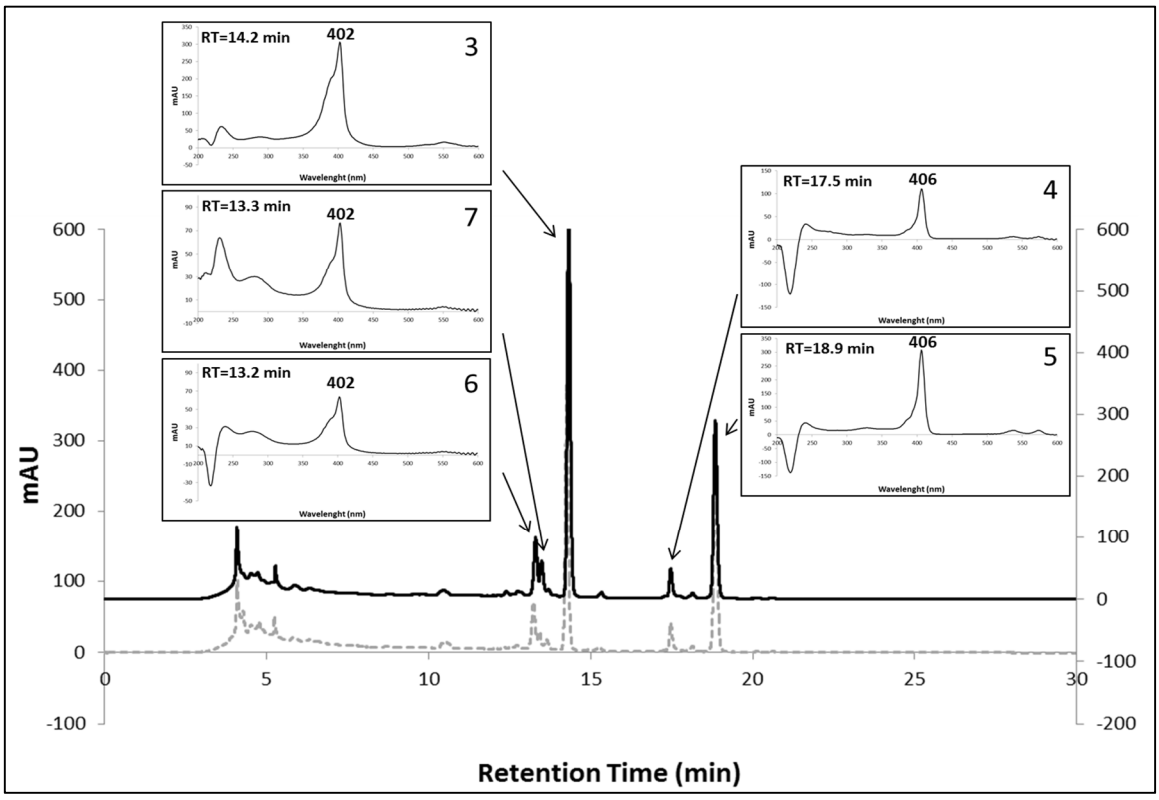

Figure 44 – HPLC chromatogram at 400 nm of *S. natalensis* Δ*pimM* [pIBSNA01239] (black line) and *S. natalensis* [pIB139] (dashed grey line) culture broth extracts. Arrows indicate peaks corresponding absorption spectra.

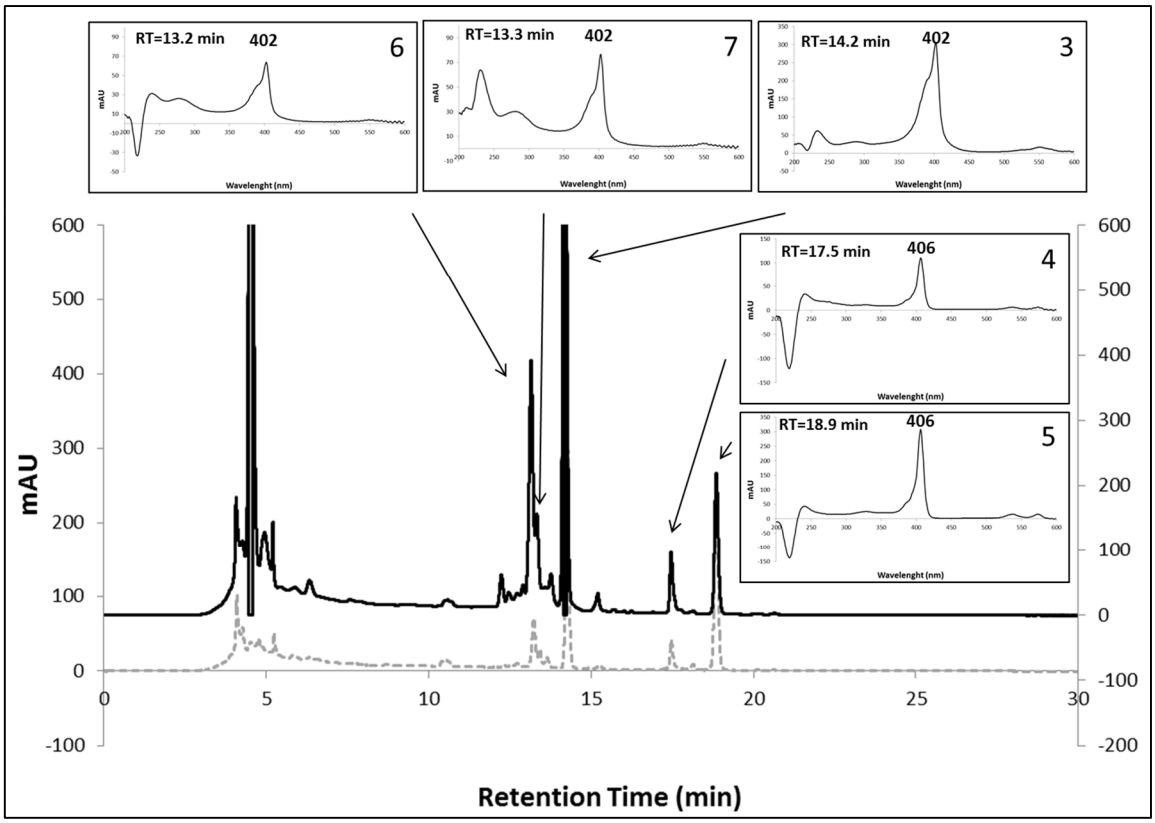

Figure 45 – HPLC chromatogram at 400 nm of *S. natalensis* Δ*pimM* [pIBSNA01173] (black line) and *S. natalensis* [pIB139] (dashed grey line) culture broth extracts. Arrows indicate peaks corresponding absorption spectra.

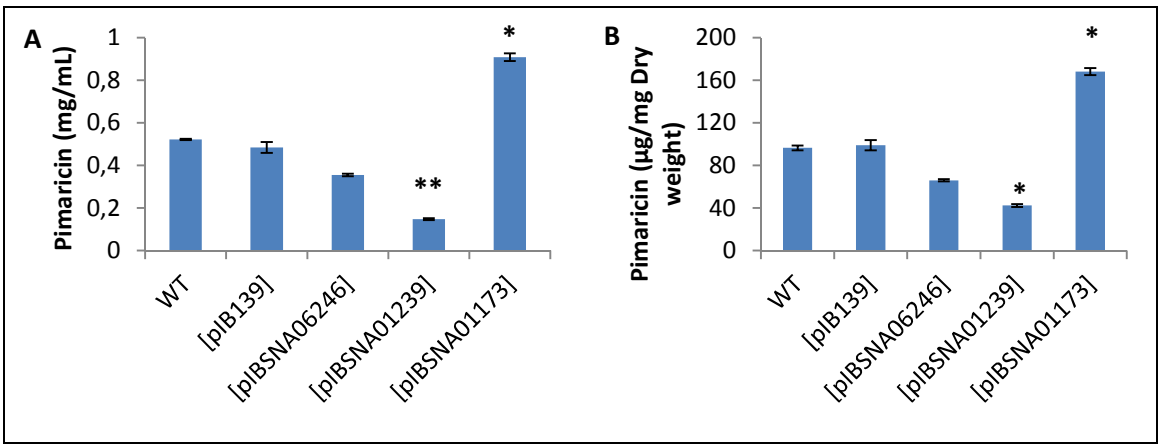

Figure 46 – Pimaricin production. A). Volumetric production. B). Specific production. The graphical results show the average obtained among two technical replicates of two independent biological replicates. Vertical bars indicate standard deviation of the mean values. The differences between exconjugants *S. natalensis* [pIBSNA01239] or *S. natalensis* [pIBSNA01239] samples and *S. natalensis* [pIB139] samples were assessed by independent t-test. (*, $P < 0.001$; **, $P < 0.05$).

3.2.5 Conclusions

The delineated strategy for overexpressing three biosynthetic gene clusters, by modulating the transcription of one of its putative positive CSR, was apparently not achieved with any of the selected ones. Laureti and its co-workers applied a similar strategy for the activation of a type I PKS from *Streptomyces ambofaciens* ATCC 23877 (Laureti *et al.*, 2011).

Though, all putative positive CSR encoding genes were overexpressed, instead of only one of them. From the three overexpressed genes, one led to the activation of the biosynthetic gene cluster. Thus, these data suggests that it should still be interesting to overexpress other CSR from the selected biosynthetic gene clusters S1C5 and S5C20, in order to activate these genomic regions.

For S1C4, as it does not hold any other regulator, this strategy could not be possible. Nevertheless, it could still be interesting to study the transcription state of one other S1C4 structural gene, since it cannot be ascertained that SNA01166 is involved in the biosynthesis of the bacteriocin. Moreover, it is known that the regulation of the biosynthesis of bacteriocins from Gram-positive bacteria relies on signal transduction systems. These compounds can be auto-regulated, in which bacteriocins activate their own expression, by interacting with a membrane-bound histidine protein kinase. Then, the internal portion of the kinase transfers a phosphoryl group to the second component, a response regulator located in the cellular cytoplasm. This causes a change in the structure of the response regulator which activates transcription. On the other hand, instead of autoregulation, a peptide pheromone induction is the one responsible for the expression of the bacteriocin. A gene unique from the structural gene encodes a bacteriocin-like substance which is the signal recognized by the extracellular domain of the histidine protein kinase (Snyder and Worobo, 2014). Since SNA01173 protein domain analysis revealed the presence of a signal receiver domain, that receives the signal from the sensor partner in a two-component system (aa 1-112), together with a C-terminal DNA-binding domain of LuxR-like proteins (aa 138-194), it is, in fact, likely to be the bacteriocin transcription activator.

Still in regard of S1C4 it is also important to mention that results did not show clearly that the studied *SNA01173* is being overexpressed in the *S. natalensis* [pIBSNA01173] replicates. RNA samples collected from *S. natalensis* [pIBSNA01173], *S. natalensis* wild-type and *S. natalensis* [pIB139] should be submitted for quantitative RT-PCR (RT-qPCR) in order to clarify the latter results.

Concerning metabolic profiles, two strains stood out: *S. natalensis* [pIBSNA01173] and *S. natalensis* [pIBSNA01239].

S. natalensis [pIBSNA01173] appears to have increased the production of the compounds represented by the chromatogram peaks #3 and #5, when compared to the control *S. natalensis* [pIB139] strain. Additionally, three novel peaks, not present in either wild-type or [pIB139] strain were also detected: #4, #6 and #7. Besides, *S. natalensis* [pIBSNA01173] showed significant increased levels of pimaricin specific production, when compared to *S. natalensis* [pIB139]. In fact, it has even reached a specific production 1.70 times higher than

the wild-type. This result, suggests that *SNA01173* is being overexpressed and, as it is a regulator from the same family as PimM (LuxR-family), this putative CSR could be exerting a pleiotropic effect and acting specifically in the pimaricin gene cluster. Therefore, the alignment of amino acid sequences of these two proteins was performed and revealed partial identity between the helix-turn-helix (HTH) regions. Similarities between both DNA binding domains could suggest similar recognized DNA binding motifs. However, the identity between these two regions is low (30%). This fact, together with no detection of pimaricin in the Δ*pimM* [pIBSNA01173] strain, can explain the fact that SNA01173 could not activate pimaricin production in Δ*pimM*. Furthermore, these data suggests a pleiotropic role for SNA01173, acting on genes unrelated and unlinked with the S1C4,including the pimaricin biosynthetics gene cluster.

In turn, it is also interesting to note that HPLC chromatogram at 400 nm of *S. natalensis* [pIBSNA01239] did not present the two peaks displayed in the control chromatogram (peaks #3 and #5). That could be due to the fact it is diverting precursors essential for the production of these metabolites, in order to produce a new one. Furthermore, *S. natalensis* [pIBSNA01239] presented a significant decrease in pimaricin specific production of 3.28 fold when compared to the control. This result is coherent with the low inhibition halos detected in the performed bioassay experiments. Worth to note that, as putative MerR-family protein, it is likely that SNA01239 acts as a transcription activator. However, it may act as a negative (and pleiotropic) regulator instead.

It is important to mention that the peaks regarding the 400 nm HPLC chromatograms appear to be by-products of the same compound due to the similarities among absorption spectra. It is not possible to know exactly which one is the end product and whose are the by-products. Therefore, it can be said that it was not detected new product in neither *S. natalensis* [pIBSNA01239], *S. natalensis* [pIBSNA06246] nor *S. natalensis* [pIBSNA01173], but only different by-products. Nevertheless, it does not ensure that new products are not being produce. Those compounds might be covered by exhibiting peaks, namely the first ones that are associated to culture medium leftovers. Additionally the selected extraction and HPLC methods could not the appropriated ones for its detection. Thus, further extraction and HPLC approaches must be performed in order to (1) gain better insights about the metabolic profile of all the developed strains and to (2) check if they whether or not produce novel bioactive compounds.

4. Concluding remarks and future perspectives

The slowdown of antimicrobial drug discovery and development, together with the continuous increase of the antimicrobial resistance phenomenon, outstands the need for novel bioactive compounds. Classical screening methods for those products have been outdated and suffer from high costs than often lead to high rediscovery rates. On the contrary, the development of high throughput technology for the screening of biological activities together with the molecular awakening of "silent" clusters may allow the unleash of secondary metabolism potential hide in the genome sequence. Additionally, synthetic biology defined as the design and construction of biological devices and systems for useful purposes (Schmidt, 2012) will also have a word to say in the discovery of novel bioactive compounds.

In this project, it was aimed to activate the transcription of specialized metabolite biosynthetic gene clusters. The activation of what is known as "silent" clusters is a promising approach to unveil novel biologically active compounds that could be further developed into hit compounds to feed the drug discovery pipeline (Olano *et al.*, 2014).

The genome mining combined with the positive CSR overexpression stands as an already proved successful and straightforward strategy for the activation of those gene clusters (Laureti *et al.*, 2011, Gottelt *et al.*, 2010). This approach was applied in the present work, by performing an in-depth *in silico S. natalensis* genome mining analysis and selecting three putative positive CSR from three independent studied silent gene clusters for overexpression. *S. natalensis* Δ*pimM*, a pimaricin non-producing strain, was also teste due to the expected precursors bioavailability resulting of no pimaricin production.

However, the gene expression results suggest that the activation of such biosynthetic gene clusters was not achieved. Several reasons may justify the latter, namely the effect of negative CSR or pleiotropic regulators, the lack of a co-factor, essential for the interaction between DNA promoter zones and the regulators, the used media or being the selected genes to act as a negative regulator s. In the particular case of S1C4 it is also conceivable that the structural gene subjected for transcription analysis does not code for a protein involved in the biosynthesis of the end product.

Nevertheless, all strains were subjected for comparative metabolic profiling analysis and interesting results were obtained with the wild-type derived strains.

The HPLC chromatograms analysis at 400 nm of *S. natalensis* [pIBSNA01173], a strain putatively overexpressing the LuxR protein encoding gene *SNA01173* from S1C4, unveiled the presence of three peaks that were not detected neither in the wild-type nor in *S. natalensis* [pIB139] strains: #4, #6 and #7. Furthermore, other two presented higher peak area per dry weight when compared to the control *S. natalensis* [pIB139] strain: #3 and #5. Still in regard of *S. natalensis* [pIBSNA01173] metabolic profile, this strain presented a significant increase in

pimaricin production. In fact, results pointed a specific production of this metabolite approximately 1.74 times higher than the wild-type strain. Therefore, SNA01173 could be involved in the activation of pimaricin biosynthetic gene cluster

Remarkable results have been also been achieved for *S. natalensis* [pIBSNA01239]. The HPLC chromatogram at 400 nm of this strain, that is overexpressing the MerR protein encoding gene *SNA01239* from S1C5, showed the disappearance of two peaks that were present in the wild-type and the in *S. natalensis* [pIB139] strains (peaks #3 and #5). Additionally, it presented significant decrease in the pimaricin specific production. This could mean that this strain is diverting the pimaricin metabolic precursors for the production of one other product.

As stated earlier, the performed extraction and HPLC methods were specially designed for pimaricin analysis. Therefore, it is possible that novel bioactive compounds are being produced by the new developed strains, although they were not detected by the performed experiments. This way, it is of great importance to perform different extraction and HPLC methods in order to look for the possible new products in a near future. By way of example, a new method of extraction could that could allow to test the whole culture broth (25 mL) instead of only 500 µL may allow the detection of compounds produced in low amounts. Using other solvents (including polar solvents, in opposition to methanol) should also be tested since the new compounds may have higher affinity to them.

Concerning the HPLC future studies, it should also be experimented to perform chromatography tests with other columns, solutions for the mobile phase and elution methods should. For example, achieve a lower percentage of the used solvent under a lower flow rate and for a longer period of time may allow unmasking certain compounds. Furthermore, in order to identify the compounds that are being produced, HPLC experiments should be repeated, but coupled to a mass spectrometer (MS) and a nuclear magnetic resonance (NMR) spectroscopy.

Carrying out new bioassay experiments using the multi-drug resistant Gram-negative bacteria *Klebsiella pneumonia*, *Acinetobacter baumannii* and *Pseudomonas aeruginosa*, as red-flagged antimicrobial priority targets (European Centre for Disease Prevention and Control, 2009, 2013), should also be performed. As stated, comparative metabolic profiling results suggest that new products might being produced, or, at least, some are being overproduced. Therefore, they may hold bioactive characteristics that could help fighting antibiotic resistance. Moreover, since cancer is a major burden of disease worldwide, accounting for 8.2 million deaths only in 2012 (Ferlay *et al*, 2013), and since anticancer compounds have already been isolated from *Streptomyces* culture (e.g. doxorubicin) it should be interesting to perform bioassays using different cancer line cells.

Once concluded the metabolic profiling of all new developed strains there are other possible future studies also of great importance. For instance, a transcription analysis throughout the growth curve instead of only at one point. This way it should be possible to get a better understanding of what genes are silenced or being expressed. Regarding S1C4 biosynthetic gene cluster, RT-qPCR may be firstly performed in order to ensure the overexpression of the putative CSR (not confirmed by the carried out RT-PCR experiments). Still in regard of S1C4, if in the latter experiments *SNA01166* did not show to be expressed in the strains putatively overexpressing *SNA01173*, one other putative structural gene should be selected. This way it can be ensured the actual silent state of the cluster, because, as mentioned, SNA01166 might not be involed in the biosynthetic pathway of the bacteriocin. *SNA01164*, as encoding for a protein with a TIGR04222 domain, typical in proteins whose C-terminal sequence resembles ribosomal natural product precursors, should be the best candidate. Additionally, since *S. natalensis* Δ*pimM* [pIBSNA01173] extract did not unveil the pimaricin peak at the 304 nm HPLC chromatogram while *S. natalensis* [pIBSNA01173] presented higher pimaricin specific production levels than the control, *pimS*1 expression should be analysed in both strains. This way, it is not regu.

If the activation of either S1C5 or S5C20 is not achieved, one other putative positive CSR should be selected to be overexpressed together with a putative negative CSR to be silenced. According to the obtained transcription data together with the *in silico* analysis, *SNA01250* from S1C5 and *SNA06239* and *SNA06243* from S5C20 are the genes that could be selected to be overexpressed. *SNA01232*, *SNA01254*, *SNA01256* and *SNA01258* from S1C5 and *SNA06240* and *SNA06245* from S5C20, due to their high transcription levels and *in silico* identification as putative negative CSRs encoding genes, could be selected to be silenced. It should be noted that the overexpression and the silencing of the proposed genes can be done separately or simultaneously, thereby increasing the chances to be succeed. In fact, the silencing of such genes may also be done in the strains overexpressing *SNA01173, SNA01239 or SNA06246*, already developed and available.

In turn, in case S1C4 is actually silent (as it does not include other putative CSR rather than SNA01173) and/or all the last gene cluster activation strategies failed to succeed, other approaches could be applied, namely the heterologous expression. *S. coelicolor* M146, a derivative strain of *S. coelicolor* M145, is a potential candidate for this approach. It lacks potentially competitive sinks of carbon and nitrogen, and, in order to provide a host devoid of antibiotic activity, its four endogenous secondary metabolite gene clusters responsible for actinorhodin, prodiginine, CPK and CDA biosynthesis are deleted. It also holds point mutations

into rpoB and rpsL to pleiotropically increase the level of secondary metabolite production (Gomez-Escribano and Bibb, 2011).

5. References

ALLENBY, N., LAING, E., BUCCA, G., KIERZEK, A. & SMITH, C. 2012. Diverse control of metabolism and other cellular processes in *Streptomyces coelicolor* by the PhoP transcription factor: genome-wide identification of in vivo targets. *Nucleic Acids Research,* 40**,** 9543-9556.

ANTÓN, N., SANTOS-ABERTURAS, J., MENDES, M. V., GUERRA, S., MARTÍN, J. & APARICIO, J. 2007. PimM, a PAS domain positive regulator of pimaricin biosynthesis in *Streptomyces natalensis*. *Microbiology,* 153**,** 3174-3183.

APARICIO, J., MENDES, M., ANTÓN, N. & MARTÍN, J. 2002. Biosynthetic rules for macrolide construction. *In:* FIERRO, F. & MARTÍN, J. (eds.) *Microbial secondary metabolites: Biosynthesis, genetics and regulation.* Trivandrum, Kerala, Indi: Research Signpost.

ARTSIMOVITCH, I., PATLAN, V., SEKINE, S., VASSYLYEVA, M., HOSAKA, T., OCHI, K., YOKOYAMA, S. & VASSYLYEV, D. 2004. Structural basis for transcription regulation by alarmone ppGpp. *Cell,* 117**,** 299-310.

AUSUBEL, F. M., BRENT, R., KINGSTON, R. E., MOORE, D. D., SEIDMAN, J. G., SMITH, J. A. & STRUHL, K. 1992. *Short protocols in molecular biology,* New York, Wiley.

BALTZ, R. 2008. Renaissance in antibacterial discovery from actinomycetes. *Current Opinion in Pharmacololy,* 8**,** 557-563.

BANGERA, M. & THOMASHOW, L. 1999. Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from *Pseudomonas fluorescens* Q2-87. *Journal of Bacteriology,* 181**,** 3155-3163.

BEASLEY, S. S. & SARIS, P. E. 2004. Nisin-producing *Lactococcus lactis* strains isolated from human milk. *Applied and Environmental Microbiology,* 70**,** 5051-5053.

BEITES, T., PIRES, S. D. S., SANTOS, C. L., OSÓRIO, H., MORADAS-FERREIRA, P. & MENDES, M. V. 2011. Crosstalk between ros homeostasis and secondary metabolism in *S. natalensis* ATCC 27448: Modulation of pimaricin production by intracellular ROS. *PLoS ONE,* 6**,** e27472.

BENTLEY, S., CHATER, K., CERDEÑO-TÁRRAGA, A., CHALLIS, G., THOMSON, N., JAMES, K., HARRIS, D., QUAIL, M., KIESER H, H. D., A, B., BROWN, S., CHANDRA, G., CHEN, C., COLLINS, M., CRONIN, A., FRASER, A., GOBLE, A., HIDALGO, J., HORNSBY, T., HOWARTH, S. & AL., E. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature,* 417**,** 141-147.

BERDY, J. 2005. Bioactive microbial metabolites. *Journal of Antibiotics (Tokyo),* 58**,** 1-26.

BERGMANN, S., SCHÜMANN, J., SCHERLACH, K., LANGE, C., BRAKHAGE, A. & HERTWECK, C. 2007. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nature Chemical Biology,* 3**,** 213-217.

BIBB, M. 2005. Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology,* 8**,** 208-215.

BLANCO, G., BRIAN, P., PEREDA, A., MENDEZ, C., SALAS, J. A. & CHATER, K. F. 1993. Hybridization and DNA sequence analyses suggest an early evolutionary divergence of related biosynthetic gene sets encoding polyketide antibiotics and spore pigments in *Streptomyces* spp. *Gene,* 130**,** 107-116.

BLIN, K., MEDEMA, M., KAZEMPOUR, D., FISCHBACH, M., BREITLING, R., TAKANO, E. & WEBER, T. 2013. antiSMASH 2.0-a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research,* 41**,** W204-W212.

BOK, J., CHIANG, Y., SZEWCZYK, E., DAVIDSON, A., SANCHEZ, J., LO, H., WATANABE, K., OAKLEY, B., WANG, C. & KELLER, N. 2009. Chromatin-level regulation of cryptic biosynthetic gene clusters in *Aspergillus nidulans*. *Nature Chemical Biology,* 5**,** 462-464.

BROWN, N., STOYANOV, J., KIDD, S. & HOBMAN, J. 2003. The MerR family of transcriptional regulators. *FEMS Microbiology Reviews,* 27**,** 145-163.

CHARUSANTI, P., FONG, N., NAGARAJAN, H., PEREIRA, A., LI, H., ABATE, E., SU, Y., GERWICK, W. & PALSSON, B. 2012. Exploiting adaptive laboratory evolution of *Streptomyces clavuligerus* for antibiotic discovery and overproduction. *PLoS One,* 7**,** e33727.

CHATER, K. & CHANDRA, G. 2006. The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiology Reviews,* 30**,** 651-672.

CHEN, J. & XIE, J. 2011. Role and regulation of bacterial LuxR-like regulators. *Journal of Cellular Biochemistry,* 112**,** 2694-702.

CHEN, X. H., VATER, J., PIEL, J., FRANKE, P., SCHOLZ, R., SCHNEIDER, K., KOUMOUTSI, A., HITZEROTH, G., GRAMMEL, N., STRITTMATTER, A. W., GOTTSCHALK, G., SUSSMUTH, R. D. & BORRISS, R. 2006. Structural and functional characterization of three polyketide synthase gene clusters *in Bacillus amyloliquefaciens* FZB 42. *Journal of Bacteriology,* 188**,** 4024-4036.

CLEVELAND, J., MONTVILLE, T. J., NES, I. F. & CHIKINDAS, M. L. 2001. Bacteriocins: safe, natural antimicrobials for food preservation. *International Journal of Food Microbiology,* 71**,** 1-20.

DAIRI, T. 2005. Studies on biosynthetic genes and enzymes of isoprenoids produced by actinomycetes. *Journal of Antibiotics (Tokyo),* 58**,** 227-243.

DEVASAHAYAM, G., SCHELD, W. M. & HOFFMAN, P. S. 2010. Newer antibacterial drugs for a new century. *Expert Opinion on Investigational Drugs,* 30**,** 215-234.

DIAS, P. J. & SA-CORREIA, I. 2013. The drug:H(+) antiporters of family 2 (DHA2), siderophore transporters (ARN) and glutathione:H(+) antiporters (GEX) have a common evolutionary origin in hemiascomycete yeasts. *BMC Genomics,* 14**,** 901.

DU, L., SANCHEZ, C. & SHEN, B. 2001. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metabolic Engineering,* 3**,** 78-95.

DURANTE-RODRIGUEZ, G., VALDERRAMA, J. A., MANCHENO, J. M., RIVAS, G., ALFONSO, C., ARIAS-PALOMO, E., LLORCA, O., GARCIA, J. L., DIAZ, E. & CARMONA, M. 2010. Biochemical characterization of the transcriptional regulator BzdR from *Azoarcus* sp. CIB. *Journal of Biological Chemistry,* 285**,** 35694-35705.

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL 2009. ECDC/ EMEA Joint Technical Report: The bacterial challenge: time to react [Online]. Available from: http://www.ecdc.europa.eu/en/publications/Publications/Forms/ECDC_DispForm.aspx?ID=444, accessed on September, 2014.

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL 2013. Annual epidemiological report 2012. Reporting on 2010 surveillance data and 2011 epidemic intelligence data. ECDC, Stockholm.

FEDOROVA, N., MOKTALI, V. & MEDEMA, M. 2012. Bioinformatics approaches and software for detection of secondary metabolic gene clusters. *Methods in Molecular Biology,* 944**,** 23-45.

FERLAY J, SOERJOMATARAM I, ERVIK M, DIKSHIT R, ESER S, MATHERS C, REBELO M, PARKIN DM, FORMAN D, BRAY, F. 2013. *GLOBOCAN 2012 v1.0, Cancer incidence and mortality worldwide: iarc cancer base no. 11* [Online]. Lyon, France: International Agency for Research on Cancer. Available from: http://globocan.iarc.fr, accessed on September 2014.

FENICAL, W. & JENSEN, P. 2006. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nature Chemical Biology,* 2**,** 666-673.

FERNÁNDEZ-MORENO, M., MARTÍN-TRIANA, A., MARTÍNEZ, E., NIEMI, J., KIESER, H., HOPWOOD, D. & MALPARTIDA, F. 1992. abaA, a new pleiotropic regulatory locus for antibiotic production in *Streptomyces coelicolor*. *Journal of Bacteriology,* 174**,** 2958-2967.

FIBRIANSAH, G., KOVÁCS, Á., POOL, T., BOONSTRA, M., KUIPERS, O. & THUNNISSEN, A. 2012. Crystal structures of two transcriptional regulators from *Bacillus cereus* define the conserved structural features of a PadR subfamily. *PLoS One,* 7**,** e48015.

FINERAN, P. C., EVERSON, L., SLATER, H. & SALMOND, G. P. 2005. A GntR family transcriptional regulator (PigT) controls gluconate-mediated repression and defines a new,

independent pathway for regulation of the tripyrrole antibiotic, prodigiosin, in *Serratia*. *Microbiology,* 151**,** 3833-2845.

FISCH, K. M. 2013. Biosynthesis of natural products by microbial iterative hybrid PKS-NRPS. *RSC Advances,* 3**,** 18228-18247.

FISCHBACH, M. & WALSH, C. 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chemical Reviews,* 106**,** 3468-3496.

FLARDH, K. & BUTTNER, M. J. 2009. *Streptomyces morphogenetics*: dissecting differentiation in a filamentous bacterium. *Nature Reviews. Microbiology,* 7**,** 36-49.

FOOD AND DRUG ADMINISTRATION 2011. Combating Antibiotic Resistance – What FDA is doing? [Online]. Available from: http://www.fda.gov/forconsumers/consumerupdates/ucm092810.htm, accessed on September 2014.

FUJII, I. 2008. Heterologous expression systems for polyketide synthases. *Natural Product Reports,* 26**,** 155-169.

FUNA, N., FUNABASHI, M., YOSHIMURA, E. & HORINOUCHI, S. 2005. A novel quinone-forming monooxygenase family involved in modification of aromatic polyketides. *The Journal of Biological Chemistry,* 280**,** 14514-14523.

FUNA, N., OHNISH, Y., EBIZUKA, Y. & HORINOUCHI, S. 2002. Properties and substrate specificity of RppA, a chalcone synthase-related polyketide synthase in *Streptomyces griseus*. *The Journal of Biological Chemistry,* 277**,** 4628-4635.

FUNA, N., OHNISHI, Y., FUJII, I., SHIBUYA, M., EBIZUKA, Y. & HORINOUCHI, S. 1999. A new pathway for polyketide synthesis in microorganisms. *Nature,* 400**,** 897-899.

GAURIVAUD, P., LAIGRET, F., GARNIER, M. & BOVE, J. M. 2000. Fructose utilization and pathogenicity of *Spiroplasma citri*: characterization of the fructose operon. *Gene,* 252**,** 61-69.

GAURIVAUD, P., LAIGRET, F., GARNIER, M. & BOVE, J. M. 2001. Characterization of FruR as a putative activator of the fructose operon of *Spiroplasma citri*. *FEMS Microbiology Letters,* 198**,** 73-78.

GOLDMAN, E., ROSENBERG, A., ZUBAY, G. & STUDIER, F. 1995. Consecutive low-usage leucine codons block translation only when near the 5'end of a message in *Escherichia coli*. *Journal of Molecular Biology,* 245**,** 467-473.

GOMEZ-ESCRIBANO, J. & BIBB, M. 2011. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters *Microbial Biotechnology,* 4**,** 207-215.

GOMEZ-ESCRIBANO, J. & BIBB, M. 2012. *Streptomyces coelicolor* as an expression host for heterologous gene clusters. *Methods in Enzymology,* 517**,** 279-300.

GOTTELT, M., KOL, S., GOMEZ-ESCRIBANO, J., BIBB, M. & TAKANO, E. 2010. Deletion of a regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). *Microbiology,* 156**,** 2343-2353.

GREENHAGEN, B. & CHAPPELL, J. 2001. Molecular scaffolds for chemical wizardry: learning nature's rules for terpene cyclases. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 13479-13481.

GROSS, H. 2007. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Applied Microbiology and Biotechnology,* 75**,** 267-277.

GRUBER, T. & GROSS, C. 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. *Review of Microbiology,* 57**,** 441-466.

GURY, J., BARTHELMEBS, L., TRAN, N., DIVIES, C. & CAVIN, J. 2004. Cloning, deletion, and characterization of PadR, the transcriptional repressor of the phenolic acid decarboxylase-encoding padA gene of *Lactobacillus plantarum*. *Applied and Environmental Microbiology,* 70**,** 2146-2153.

GUST, B., CHANDRA, G., JAKIMOWICZ, D., YUQING, T., BRUTON, C. & CHATER, K. 2004. Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Advances in Applied Microbiology,* 54**,** 107-128.

HARDISSON, C., MANZANAL, M.-B., SALAS, J.-A. & SUAREZ, J.-E. 1978. Fine structure, physiology and biochemistry of arthrospore germination in *Streptomyces antibioticus*. *Journal of General Microbiology,* 105**,** 203-214.

HIGGENS, C. E., HAMILL, R. L., SANDS, T. H., HOEHN, M. M. & DAVIS, N. E. 1974. Letter: The occurrence of deacetoxycephalosporin C in fungi and streptomycetes. *Journal of Antibiotics (Tokyo),* 27**,** 298-300.

HILLERICH, B. & WESTPHELING, J. 2006. A new GntR family transcriptional regulator in *Streptomyces coelicolor* is required for morphogenesis and antibiotic production and controls transcription of an ABC transporter in response to carbon source. *Journal of Bacteriology,* 188**,** 7477-7487.

HOBBS, G., CM, F., GARDNER, D., CULLUM, J. & OLIVER, S. 1989. Dispersed growth of *Streptomyces* in liquid culture. *Applied Microbiology and Biotechnology,* 31**,** 272-277.

HORINOUCHI, S., KITO, M., NISHIYAMA, M., FURUYA, K., HONG, S. K., MIYAKE, K. & BEPPU, T. 1990. Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2). *Gene,* 95**,** 49-56.

HOSAKA, T., OHNISHI-KAMEYAMA, M., MURAMATSU, H., MURAKAMI, K., TSURUMI, Y., KODANI, S., YOSHIDA, M., FUJIE, A. & OCHI, K. 2009. Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nature Biotechnology,* 27**,** 462-464.

HUANG, J., SHI, J., MOLLE, V., SOHLBERG, B., WEAVER, D., BIBB, M., KAROONUTHAISIRI, N., LIH, C.-J., KAO, C., BUTTNER, M. & COHEN, S. 2005. Crossregulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Molecular Microbiology,* 58**,** 1276-1287.

HWANG, K.-S., KIMA, H., CHARUSANTI, P., PALSSON, B. & LEE, S. 2014. Systems biology and biotechnology of *Streptomyces* species for the. *Biotechnology Advances,* 32**,** 255-268.

JONES, A., GUST, B., KULIK, A., HEIDE, L., BUTTNER, M. & BIBB, M. 2013. Phage p1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. *PLoS One,* 8**,** e69319.

KANE, J. 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology,* 6**,** 494-500.

KARPIŃSKI, T. & SZKARADKIEWICZ, A. 2013. Characteristic of bacteriocines and their application. *Polish Journal of Microbiology,* 62**,** 223-235.

KELEMEN, G. H., BRIAN, P., FLARDH, K., CHAMBERLIN, L., CHATER, K. F. & BUTTNER, M. J. 1998. Developmental regulation of transcription of *whiE*, a locus specifying the polyketide spore pigment in *Streptomyces coelicolor* A3 (2). *J Bacteriol,* 180**,** 2515-25121.

KHALDI, N., SEIFUDDIN, F., TURNER, G., HAFT, D., NIERMAN, W., WOLFE, K. & FEDOROVA, N. 2010. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology,* 47**,** 736-741.

KIESER, T., BIBB, M., BUTTNER, M., CHATER, K. & HOPWOOD, D. 2000. *Practical Streptomyces genetics*, Norwich John Innes Foundation.

KOLB, A., BUSBY, S., BUC, H., GARGES, S. & ADHYA, S. 1993. Transcriptional regulation by cAMP and its receptor protein. *Annual Reviews of Biochemistry,* 62**,** 749-795.

KOMATSU, M., TSUDA, M., ŌMURA, S., OIKAWA, S. & IKEDA, H. 2008. Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. *Proceedings of the National Academy of Sciences,* 105**,** 7422-7427.

KOMATSU, M., UCHIYAMA, T., OMURA, S., CANE, D. & IKEDA, H. 2010. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proceedings of the National Academy of Sciences,* 107**,** 2646-2651.

KORNER, H., SOFIA, H. J. & ZUMFT, W. G. 2003. Phylogeny of the bacterial superfamily of CRP-FNR transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiology Reviews,* 27**,** 559-592.

LAI, C., XU, J., TOZAWA, Y., OKAMOTO-HOSOYA, Y., YAO, X. & OCHI, K. 2002. Genetic and physiological characterization of rpoB mutations that activate antibiotic production in *Streptomyces lividans*. *Microbiology,* 148**,** 3365-3373.

LAURETI, L., SONG, L., HUANG, S., CORRE, C., LEBLOND, P., CHALLIS, G. & AIGLE, B. 2011. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 6258-6263.

LEE, P. C., UMEYAMA, T. & HORINOUCHI, S. 2002. AfsS is a target of AfsR, a transcriptional factor with ATPase activity that globally controls secondary metabolism in *Streptomyces coelicolor* A3(2). *Molecular Microbiology,* 43**,** 1413-1430.

LIN, X., HOPSON, R. & CANE, D. 2006. Genome mining in Streptomyces coelicolor: molecular cloning and characterization of a new sesquiterpene synthase. *Journal of the American Chemical Society,* 128**,** 6022-6023.

LIU, G., CHATER, C., CHANDRA, G., NIU, G. & TAN, H. 2013. Molecular Regulation of Antibiotic Biosynthesis in *Streptomyces*. *Microbiology and Molecular Biology Reviews,* 77**,** 112-143.

MACNEIL, D., GEWAIN, KM, RUBY, CL, DEZENY, G, GIBBONS, PH, MACNEIL, T 1992. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene,* 111**,** 61-68.

MAKITRYNSKYY, R., OSTASH, B., TSYPIK, O., REBETS, Y., DOUD, E., MEREDITH, T., LUZHETSKYY, A., BECHTHOLD, A., WALKER, S. & FEDORENKO, V. 2013. Pleiotropic regulatory genes *bldA*, *adpA* and *absB* are implicated in production of phosphoglycolipid antibiotic moenomycin. *Open Biology,* 3**,** 130121.

MARAHIEL, M. & ESSEN, L. 2009. Chapter 13 nonribosomal peptide synthetases: mechanistic and structural aspects of essential domains. *Methods in Enzymology,* 459**,** 337-351.

MARCHLER-BAUER, A., ZHENG, C., CHITSAZ, F., DERBYSHIRE, M. K., GEER, L. Y., GEER, R. C., GONZALES, N. R., GWADZ, M., HURWITZ, D. I., LANCZYCKI, C. J., LU, F., LU, S., MARCHLER, G. H., SONG, J. S., THANKI, N., YAMASHITA, R. A., ZHANG, D. & BRYANT, S. H. 2013. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res,* 41**,** D348-D352.

MARTIN, R. & ROSNER, J. 2001. The AraC transcriptional activators. *Current Opinion in Microbiology,* 4**,** 132-137.

MARVAUD, J. C., EISEL, U., BINZ, T., NIEMANN, H. & POPOFF, M. R. 1998. TetR is a positive regulator of the tetanus toxin gene in *Clostridium tetani* and is homologous to *botR*. *Infect Immun,* 66**,** 5698-702.

MCINTOSH, J. A., DONIA, M. S. & SCHMIDT, E. W. 2009. Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. *Nat Prod Rep,* 26**,** 537-59.

MEDEMA, M., BLIN, K., CIMERMANCIC, P., DE JAGER, V., ZAKRZEWSKI, P., FISCHBACH, M., WEBER, T., TAKANO, E. & BREITLING, R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research,* 39**,** W339-W346.

MENDES, M. V., ANTON, N., MARTIN, J. F. & APARICIO, J. F. 2005. Characterization of the polyene macrolide P450 epoxidase from *Streptomyces natalensis* that converts de-epoxypimaricin into pimaricin. *Biochemical Journal,* 386**,** 57-62.

MILLER, J. 1972. *Experiments in molecular genetics,* New York, Cold Spring Harbor Laboratory Press, U.S.

MOORE, B. 2008. Extending the biosynthetic repertoire in ribosomal peptide assembly. *Angewandte Chemie International Edition,* 47**,** 9386-9388.

MOORE, B. & HERTWECK, C. 2002. Biosynthesis and attachment of novel bacterial polyketide synthase. *Natural Product Reports,* 19**,** 70-99.

NETT, M., IKEDA, H. & MOORE, B. 2009. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Natural Product Reports,* 26**,** 1362-1384.

NEWCOMBE, G. & COOK, D. 2002. Influences on the removal of tastes and odours by PAC. *Journal of Water Supply: Research and Technology - Aqua,* 51**,** 463-474

NIESELT, K., BATTKE, F., HERBIG, A., BRUHEIM, P., WENTZEL, A., JAKOBSEN, O. M., SLETTA, H., ALAM, M. T., MERLO, M. E., MOORE, J., OMARA, W. A., MORRISSEY, E. R., JUAREZ-HERMOSILLO, M. A., RODRIGUEZ-GARCIA, A., NENTWICH, M., THOMAS, L., IQBAL, M., LEGAIE, R., GAZE, W. H., CHALLIS, G. L., JANSEN, R. C., DIJKHUIZEN, L., RAND, D. A., WILD, D. L., BONIN, M., REUTHER, J., WOHLLEBEN, W., SMITH, M. C., BURROUGHS, N. J., MARTIN, J. F., HODGSON, D. A., TAKANO, E., BREITLING, R., ELLINGSEN, T. E. & WELLINGTON, E. M. 2010. The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics,* 11**,** 10.

NIEWERTH, H., PARSCHAT, K., RAUSCHENBERG, M., RAVOO, B. & FETZNER, S. 2013. The PaaX-type repressor MeqR2 of *Arthrobacter* sp. strain Rue61a, involved in the regulation of quinaldine catabolism, binds to its own promoter and to catabolic promoters and specifically responds to anthraniloyl coenzyme A. *Journal of Bacteriology,* 195**,** 1068-16080.

NIGAM, A., GUPTA, D. & SHARMA, A. 2014. Treatment of infectious disease: Beyond antibiotics. *Microbiological Research*. 169**,** 643-651.

OCHI, K. 2007. From microbial differentiation to ribosome engineering. *Bioscience, Biotechnology, and Biochemistry,* 71**,** 1373-1386.

OCHOA, S. 1951. Biological mechanisms of carboxylation and decarboxylation. *Physiological Reviews,* 31**,** 56-106.

OLANO, C., GARCÍA, I., GONZÁLEZ, A., RODRIGUEZ, M., ROZAS, D., RUBIO, J., SÁNCHEZ-HIDALGO, M., BRAÑA, A. F., MÉNDEZ, C. & SALAS, J. A. 2014. Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microbial Biotechnology,* 7**,** 242-256.

OSTASH, B., SAGHATELIAN, A. & WALKER, S. 2007. A streamlined metabolic pathway for the biosynthesis of moenomycin A. *Chemistry & Biology,* 14**,** 257-267.

PAGET, M., CHAMBERLIN, L., ATRIH, A., FOSTER, S. & BUTTNER, M. 1999. Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *Journal of Bacteriology,* 181**,** 204-211.

PÉREZ-RUEDA, E. & COLLADO-VIDES, J. 2001. Common history at the origin of the position-function correlation in transcriptional regulators in archaea and bacteria. *Journal of Molecular Evolution,* 53**,** 172-179.

PFEIFER, V., NICHOLSON, G., RIES, J., RECKTENWALD, J., SCHEFER, A., SHAWKY, R., SCHRÖDER, J., WOHLLEBEN, W. & PELZER, S. 2001. A polyketide synthase in glycopeptide biosynthesis - the biosynthesis of the non-proteinogenic amino acid (s)-3,5-dihydroxyphenylglycine. *The Journal of Biological Chemistry,* 276**,** 38370-38377.

RACHID, S., REVERMANN, O., DAUTH, C., KAZMAIER, U. & MULLER, R. 2010. Characterization of a novel type of oxidative decarboxylase involved in the biosynthesis of the styryl moiety of chondrochloren from an acylated tyrosine. *Journal of Biological Chemistry,* 285**,** 12482-12489.

RAMOS, J., MARTÍNEZ-BUENO, M., MOLINA-HENARES, A., TERÁN, W., WATANABE, K., ZHANG, X., GALLEGOS, M., BRENNAN, R. & TOBES, R. 2005. The TetR family of transcriptional repressors. *Microbiology and Molecular Biology Reviews,* 69**,** 326-356.

RAWLINGS, N. D. & BARRETT, A. J. 1994. Families of serine peptidases. *Methods in Enzymology,* 244**,** 19-61.

RIX, U., FISCHER, C., REMSING, L. & ROHR, J. 2002. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Natural Product Reports,* 19**,** 542-580.

SAMBROOK, J. & RUSSEL, D. 2001. *Molecular cloning: A laboratory manual,* New York, Cold Spring Harbor Laboratory Press.

SANTOS-ABERTURAS, J., VICENTE, C. M., GUERRA, S. M., PAYERO, T. D., MARTIN, J. F. & APARICIO, J. F. 2011. Molecular control of polyene macrolide biosynthesis: direct binding of the regulator PimM to eight promoters of pimaricin genes and identification of binding boxes. *Journal of Biological Chemistry,* 286**,** 9150-9161.

SANTOS-BENEIT, F., RODRIGUEZ-GARCIA, A., SOLA-LANDA, A. & MARTIN, J. 2009. Cross-talk between two global regulators in *Streptomyces*: PhoP and AfsR interact in the control of *afsS*, *pstS* and *phoRP* transcription. *Molecular Microbiology,* 72**,** 53-68.

SCHEFFLER, R., COLMER, S., TYNAN, H., DEMAIN, A. & GULLO, V. 2013. Antimicrobials, drug discovery, and genome mining. *Applied Microbiology and Biotechnology,* 97**,** 969-978.

SCHELL, M. 1993. Molecular biology of the LysR family of transcriptional regulator. *Annual Review of Microbiology,* 47**,** 597-626.

SCHERLACH, K. & HERTWECK, C. 2006. Discovery of aspoquinolones A-D, prenylated quinoline-2-one alkaloids from *Aspergillus nidulans*, motivated by genome mining. *Organical & Biomolecular Chemistry,* 4**,** 3517-3520.

SCHERLACH, K. & HERTWECK, C. 2009. Triggering cryptic natural product biosynthesis in microorganisms. *Organical & Biomolecular Chemistry,* 7**,** 1753-1760.

SCHMIDT, A., WÄCHTLER, B., TEMP, U., KREKLING, T., SÉGUIN, A. & GERSHENZON, J. 2010. A bifunctional geranyl and geranylgeranyl diphosphate synthase is involved in terpene oleoresin formation in *Picea abies*. *Plant Physiology,* 152**,** 639-655.

SCHMIDT, M. 2012. *Synthetic biology industrial and environmental applications* [Online]. Weinheim: Wiley-Blackwell. Available from: http://site.ebrary.com/id/10565131, accessed on September 2014.

SCHROECKH, V., SCHERLACH, K., NÜTZMANN, H.W., SHELEST, E., SCHMIDT-HECK, W, SCHUEMANN, J., MARTIN, K., HERTWECK, C. & BRAKHAGE, A. 2009. Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proceedings of the National Academy of Sciences of the United States of America,* 34**,** 14558-14563.

SHIMA, J., HESKETH, A., OKAMOTO, S., KAWAMOTO, S. & OCHI, K. 1996. Induction of actinorhodin production by rpsL (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *Journal of Bacteriology,* 178**,** 7276–7284.

SIEBER, S. & MARAHIEL, M. 2005. Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chemical Reviews,* 105**,** 715−738.

SKERLOVÁ, J., FÁBRY, M., HUBÁLEK, M., OTWINOWSKI, Z. & REZÁČOVÁ, P. 2014. Structure of the effector-binding domain of deoxyribonucleoside regulator DeoR from *Bacillus subtilis*. *FEBS Journal*. 281**,** 4280-4292.

SNYDER, A. B. & WOROBO, R. W. 2014. Chemical and genetic characterization of bacteriocins: antimicrobial peptides for food safety. *Journal of the Science of Food and Agriculture,* 94**,** 28-44.

SOISSON, S., MACDOUGALL-SHACKLETON, B., SCHLEIF, R. & WOLBERGER, C. 1997. Structural basis for ligand-regulated oligomerization of AraC. *Science,* 276**,** 421-425.

SONG, L., BARONA-GOMEZ, F., CORRE, C., XIANG, L., UDWARY, D., AUSTIN, M., NOEL, J., MOORE, B. & CHALLIS, G. 2006. Type III polyketide synthase β-ketoacyl-ACP starter unit and ethylmalonyl-coa extender unit selectivity discovered by *Streptomyces coelicolor* genome mining. *American Chemical Society,* 128**,** 14754-14755.

STAUNTON, J. & WEISSMAN, K. 2001. Polyketide biosynthesis: a millennium review. *Natural Products Reports,* 18**,** 380-416.

TERPE, K. 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology,* 72**,** 211-222.

THOMAS, M. C. & CHIANG, C. M. 2006. The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology,* 41**,** 105-178.

WALSH, C. & FISCHBACH, M. 2010. Natural products version 2.0: connecting genes to molecules. *Journal of the American Chemical Society,* 132**,** 2469-2493.

WANG, R., MAST, Y., WANG, J., ZHANG, W., ZHAO, G., WOHLLEBEN, W., LU, Y. & JIANG, W. 2012. Identification of two-component system AfsQ1/Q2 regulon and its cross-regulation with GlnR in *Streptomyces coelicolor*. *Molecular Microbiology,* 87**,** 30-48.

WATVE, M. G., TICKOO, R. & JOG, M. M. 2001. How many antibiotics are produced by the genus *Streptomyces*? *Archives of Microbiology,* 176**,** 386-390.

WENZEL, S. & MÜLLER, R. 2009. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Natural Products Reports,* 26**,** 1385-1407.

WILKINSON, C., HUGHES-THOMAS, Z., MARTIN, C., BOHM, I., MIRONENKO, T., DEACON, M., WHEATCROFT, M., WIRTZ, G., STAUNTON, J. & LEADLAY, P. 2002. Increasing the efficiency of heterologous promoters in actinomycetes. *Journal of Molecular Microbiology and Biotechnology,* 4**,** 417-426.

WILLIAMS, R., HENRIKSON, J., HOOVER, A., LEE, A. & CICHEWICZ, R. 2008. Epigenetic remodeling of the fungal secondary metabolome. *Organical & Biomolecular Chemistry,* 6**,** 1895-1897.

XU, J., TOZAWA, Y., LAI, C., HAYASHI, H. & OCHI, K. 2002. A rifampicin resistance mutation in therpoB gene confers ppGpp-independent antibiotic production in *Streptomyces coelicolor* A3(2). *Molecular Genetics and Genomics,* 268**,** 179-189.

YAGUE, P., LOPEZ-GARCIA, M. T., RIOSERAS, B., SANCHEZ, J. & MANTECA, A. 2012. New insights on the development of and their relationships with secondary metabolite production. *Current Trends in Microbiology,* 8**,** 65-73.

YU, D., XU, F., ZENG, J. & ZHAN, J. 2012. Type III polyketide synthases in natural product biosynthesis. *International Union of Biochemistry and Molecular Biology Life,* 64**,** 285-295.

ZENG, J., DECKER, R. & ZHAN, J. 2012. Biochemical Characterization of a Type III polyketide biosynthetic gene cluster from *Streptomyces toxytricini*. *Applied Biochemistry and Biotechnology,* 166**,** 1020-1033.

ZHANG, H., CHEN, J., WANGA, H., XIE, Y., JU, J., YAN, Y. & ZHANG, H. 2013. Structural analysis of HmtT and HmtN involved in the tailoring steps of himastatin biosynthesis. *FEBS Letters,* 587**,** 1675-1680.

6. Appendix I

Table A1 – BLASTp analysis of S1C1 encoding genes

Gene	Best hit description	E-value (% Identity)
SNA00084	Hypothetical protein (*Streptomyces* sp. BoleA5)	1e-41 (39%)
SNA00085	Hypothetical protein (*Streptomyces* sp. HmicA12)	2e-10 (44%)
SNA00086	Hypothetical protein (*Streptomyces* sp. HmicA12)	7e-81 (56%)
SNA00087	Hypothetical protein (*Streptomyces* sp. MspMP-M5)	1e-169 (77%)
SNA00088	LysR family transcriptional regulator (*Streptomyces albulus*)	5e-150 (81%)
SNA00089	Hypothetical protein (*Streptomyces* sp. MspMP-M5)	5e-143 (80%)
SNA00090	Uncharacterized protein UHOR_00132 (*Ustilago hordei*)	0.93 (50%)
SNA00091	Hypothetical protein (*Streptomyces* sp. MspMP-M5)	1e-34(93%)
SNA00092	Glyoxalase (*Streptomyces venezuelae*)	1e-65 (85%)
SNA00094	SAM-dependent methlyltransferase (*Streptomyces* sp. NRRL S-237)	0.0 (88%)
SNA00095	Terpene synthase metal-binding domain-containing protein (*Streptomyces violaceusniger* Tu 4113)	0.0 (82%)
SNA00096	Crp/Fnr family transcriptional regulator (*Streptomyces* sp. NRRL S-237)	1e-21 (73%)
SNA00097	Hypothetical protein (*Streptomyces auratus*)	3e-36 (83%)
SNA00098	Hypothetical protein (*Streptomyces auratus*)	1e-124 (93%)
SNA00100	AMP-dependent synthetase (*Streptomyces filamentosus*)	2e-167 (56%)
SNA00101	Phosphopantetheine-binding protein (*Streptomyces roseochromogenes*)	1e-21 (60%)
SNA00102	Decarboxylase (*Actinoplanes* sp. SE50/110)	4e-129 (54%)
SNA00103	Hypothetical protein (*Streptomyces roseochromogenesi*)	2e-10 (29%)
SNA00104	Hypothetical protein (*Streptomyces albulus*)	1e-19 (38%)
SNA00105	Hypothetical protein BMG523Draft_00765 (*Frankia* sp. BMG5.23)	3e-75 (63%)

Table A2 – BLASTp analysis of S1C2 encoding genes

Gene	Best hit description	E-value (% Identity)
SNA00374	Acyl-CoA dehydrogenase (Streptomyces sp. FxanaC1)	0.0 (85%)
SNA00375	Histidine kinase (*Streptomyces auratus*)	3e-120 (85%)
SNA00376	ATP-binding protein (*Streptomyces auratus*)	2e-120 (80%)
SNA00377	Hypothetical protein (*Streptomyces* sp. FxanaC1)	8e-75(90%)
SNA00378	Dynein regulation protein LC7 (*Streptomyces* sp. FxanaC1)	9e-77 (75%)
SNA00379	Sensor histidine kinase, partial (*Streptomyces rimosus*)	2e-154 (69%)
SNA00381	Hypothetical protein (*Streptomyces aurantiacus*)	5e-89 (69%)
SNA00382	Hypothetical protein (*Streptomyces aurantiacus*)	4e-13 (61%)
SNA00383	Metallo-beta-lactamase (*Streptomyces prunicolor*)	0.0 (69%)
SNA00384	Hypothetical protein (*Streptomyces* sp. MspMP-M5)	0.0 (79%)
SNA00385	Hypothetical protein (*Streptomyces* sp. MspMP-M5)	0.0 (74%)
SNA00386	Hypothetical protein (*Streptomyces thermolilacinus*)	0.0 (58%)
SNA00387	Hypothetical protein (*Streptomyces purpureus*)	6e-100 (50%)
SNA00388	Hypothetical protein (*Streptomyces* sp. MspMP-M5)	1e-124 (66%)
SNA00389	Endonuclease (*Streptomyces violaceusniger*)	4e-143 (73%)
SNA00390	C14 family peptidase (*Streptomyces* sp. SS)	9e-130 (65%)
SNA00391	Hypothetical protein (*Streptomyces* sp. FxanaC1)	2e-22 (73%)
SNA00392	Secreted protein (*Streptomyces auratus* AGR0001)	9e-129 (56%)
SNA00393	O-methyltransferase family 2 (*Streptomyces albulus*)	0.0 (82%)
SNA00394	Polyketide cyclase WhiE VII (*Streptomyces venezuelae* ATCC 10712)	4e-58 (81%)
SNA00395	Aromatase WhiE VI (*Streptomyces albulus*)	7e-89 (78%)
SNA00396	Acyl carrier protein (*Streptomyces auratus* AGR0001)	5e-35 (73%)
SNA00397	Polyketide chain length factor WhiE-CLF (*Streptomyces gancidicus*)	0.0 (78%)
SNA00398	Polyketide beta-ketoacyl synthase WhiE-KS (*Streptomyces albulus*)	0.0 (89%)
SNA00399	Polyketide cyclase WhiE II (*Streptomyces gancidicus* BKS 13-15)	8e-67 (72%)
SNA00400	SchA/CurD domain-containing protein (*Streptomyces albulus*)	0.0 (78%)
SNA00401	Monooxygenase FAD-binding protein (*Streptomyces albulus*)	0.0 (79%)
SNA00402	Hypothetical protein (*Streptomyces* sp. FxanaC1)	7e-151 (71%)
SNA00403	AbaA-like protein (*Streptomyces albulus*)	1e-52 (93%)
SNA00404	DNA-binding protein (XRE-family like protein) (*Streptomyces* sp. FxanaC1)	0.0 (93%)
SNA00405	ATP-binding protein (*Streptomyces albulus*)	7e-67 (84%)
SNA00406	SGNH hydrolase (*Streptomyces auratus*)	6e-174 (91%)
SNA00407	Protein tyrosine/serine phosphatase (*Streptomyces albulus* CCRC 11814)	2e-159 (85%)
SNA00408	aromatic ring-opening dioxygenase LigA (*Streptomyces bingchenggensis*)	8e-131 (59%)
SNA00409	Aminotransferase (*Streptomyces rimosus*)	0.0 (89%)
SNA00410	1-deoxy-D-xylulose-5-phosphate synthase (*Streptomyces auratus*)	0.0 (88%)
SNA00411	Radical SAM protein (*Streptomyces auratus*)	0.0 (96%)
SNA00412	Hypothetical protein (*Streptomyces rimosus*)	8e-99 (79%)
SNA00413	Squalene-hopene cyclase (*Streptomyces* sp. FxanaC1)	0.0 (84%)
SNA00414	Dimethylallyltranstransferase (*Streptomyces* sp. MspMP-M5)	0.0 (94%)
SNA00415	Squalene/phytoene dehydrogenase (*Streptomyces auratus* AGR0001)	0.0 (85%)

Table A2 – BLASTp analysis of S1C2 encoding genes (continuation)

Gene	Best hit description	E-value (% Identity)
SNA00416	Phytoene synthase (*Streptomyces* sp. FxanaC1)	0.0 (91%)
SNA00417	Squalene synthase (*Streptomyces* sp. FxanaC1)	6e-175 (91%)
SNA00418	Hypothetical protein (*Candidatus Latescibacter anaerobius*)	2.0 (47%)
SNA00419	Techoic acid ABC transporter ATP-binding protein (*Streptomyces* sp. MspMP-M5)	0.0 (93%)
SNA00420	ABC transporter (*Streptomyces auratus*)	0.0 (90%)
SNA00421	Glycosyl transferase (*Streptomyces prunicolor*)	5e-177 (86%)
SNA00422	Transferase (*Streptomyces* sp. FxanaC1)	1e-141 (90%)
SNA00423	Dehydrogenase (*Streptomyces auratus*)	0.0 (95%)
SNA00424	Putative mannose-1-phosphate guanylyltransferase 1 (*Streptomyces aurantiacus*)	5e-171 (98%)
SNA00425	Transferase (*Streptomyces* sp. FxanaC1)	0.0 (86%)
SNA00426	Multidrug resistance-associated protein 6 (*Canis lupus familiaris*)	6.2 (38%)
SNA00427	Isopentenyl-diphosphate delta-isomerase (*Streptomyces albulus*)	2e-126 (92%)
SNA00428	Regulatory protein (*Streptomyces auratus*)	9e-72 (86%)
SNA00429	Hypothetical protein CARUB_v10012271mg (*Capsella rubella*)	1.6 (46%)
SNA00430	Enoyl-CoA hydratase (*Streptomyces* sp. FxanaC1)	0.0 (89%)
SNA00431	Heme ABC transporter ATP-binding protein (*Streptomyces* sp. FxanaC1)	1e-147 (92%)
SNA00432	Glutamine amidotransferase (*Streptomyces auratus*)	2e-124 (84%)
SNA00433	AraC family transcriptional regulator (*Streptomyces* sp. FxanaC1)	7e-179 (85%)
SNA00434	Hypothetical protein (*Streptomyces* sp. FxanaC1)	6e-164 (83%)
SNA00435	RNA-binding protein (*Streptomyces* sp. FxanaC1)	0.0 (92%)
SNA00436	Hypothetical protein (*Streptomyces* sp. FxanaC1)	0.0 (87%)
SNA00437	ATPase AAA (*Streptomyces niveus*)	0.0 (69%)

Table A3 – BLASTp analysis of S1C3 encoding genes

Gene	Best hit description	E-value (% Identity)
SNA00822	Transposase (*Mycobacterium avium* subsp. hominissuis TH135)	6e-136 (51%)
SNA00823	Uracil-DNA glycosylase (*Rhizobium grahamiil*)	6.9 (32%)
SNA00824	Hypothetical protein (*Streptomyces cattleyal*)	0.0 (66%)
SNA00825	Hypothetical protein (*Streptomyces canus*)	0.015 (77%)
SNA00826	IS1647-like transposase (*Streptomyces* sp. GBA 94-10)	9e-71 (85%)
SNA00827	Glyoxalase (*Streptomyces auratus*)	1e-141 (77%)
SNA00828	Hypothetical protein (*Streptomyces auratus*)	0.0 (84%)
SNA00829	Hypothetical protein (*Streptomyces auratus*)	3e-64 (73%)
SNA00830	Regulator (*Streptomyces filamentosus*)	7e-165 (48%)
SNA00831	SAM-dependent methyltransferase (*Streptomyces auratus*	1e-126 (70%)
SNA00832	Hypothetical protein (*Streptomyces auratus*)	1e-178 (75%)
SNA00833	Ribonuclease BN (*Streptomyces* sp. FxanaC1)	2e-158 (84%)
SNA00834	Pterin-4-alpha-carbinolamine dehydratase (*Streptomyces* sp. FxanaC1)	2e-53 (81%)
SNA00835	XRE family transcriptional regulator (*Streptomyces* sp. FxanaC1)	3e-161 (88%)
SNA00836	Isomerase (*Streptomyces* sp. PsTaAH-124)	1e-30 (74%)
SNA00837	Hypothetical protein (*Streptomyces* sp. CNT372)	1e-53 (77%)
SNA00838	Transporter (*Streptomyces auratus*)	0.0 (90%)
SNA00839	Hypothetical protein (*Streptomyces* sp. FxanaC1	0.0 (78%)
SNA00840	Aminopeptidase (*Streptomyces auratus*)	0.0 (77%)
SNA00841	Type III polyketide synthase RppA (*Streptomyces albulus*)	0.0 (89%)
SNA00842	Hypothetical protein KGM_06486 (*Danaus plexippus*)	1.5 (28%)
SNA00843	Cytochrome P450 (*Streptomyces albulus*)	0.0 (84%)
SNA00844	Phosphoesterase (*Streptomyces albulus*)	0.0 (78%)
SNA00845	UbiE family methyltransferase (*Streptomyces auratus*)	2e-170 (87%)
SNA00846	Glyoxalase (*Streptomyces* sp. MspMP-M5)	1e-71 (96%)
SNA00847	Alcohol dehydrogenase (*Streptomyces auratus*)	0.0 (83%)
SNA00848	Hypothetical protein (*Streptomyces* sp. FxanaC1)	0.0 (82%)
SNA00849	Acetyltransferase (*Streptomyces* sp. FxanaC1)	1e-94 (76%)
SNA00850	Transposase (*Streptomyces fradiae*)	0.0 (77%)
SNA00851	TetR family transcriptional regulator (*Streptomyces sp*. FxanaC1)	3e-92 (77%)
SNA00852	Hypothetical protein (*Streptomyces* sp. FxanaC1)	4e-137 (70%)
SNA00853	Hypothetical protein (*Streptomyces* sp. FxanaC1)	4e-137 (70%)
SNA00854	Ferredoxin (*Streptomyces tubercidicus*)	1e-29 (84%)
SNA00855	S15 family peptidase (*Streptomyces albulus*)	0.0 (79%)
SNA00856	Acyl-peptide hydrolase (*Streptomyces* sp. FxanaC1)	0.0 (84%)
SNA00858	Hypothetical protein (*Streptomyces albulus*)	0.0 (86%)
SNA00859	Peptide ABC transporter substrate-binding protein (*Streptomyces* sp. FxanaC1)	0.0 (94%)

Table A4 – BLASTp analysis of S1C4 encoding genes

Gene	Best hit description	E-value (% Identity)
SNA01156	ATPase (*Streptomyces* sp. FxanaC1)	0.0 (90%)
SNA01157	Predicted protein (*Streptomyces iranensis*)	2e-19 (95%)
SNA01158	ABC transporter (*Streptomyces rimosus*)	4e-77 (72%)
SNA01159	ABC transporter (*Streptomyces* sp. FxanaC1)	3e-151 (76%)
SNA01160	Peptidyl-tRNA hydrolase (*Streptomyces auratus*)	3e-147 (77%)
SNA01161	Hypothetical protein (*Streptomyces* sp. FxanaC1)	2e-125 (89%)
SNA01162	Endonuclease (*Streptomyces rapamycinicus*)	3e-166 (89%)
SNA01163	Endonuclease (*Streptomyces purpureus*)	7e-52 (65%)
SNA01164	Hypothetical protein (*Streptomyces albulus*)	6e-68 (50%)
SNA01165	Hypothetical protein (*Streptomyces* sp. MspMP-M5)	3e-76 (60%)
SNA01166	Tripeptidyl aminopeptidase (*Streptomyces auratus*)	0.0 (71%)
SNA01167	Hypothetical protein (*Streptomyces* sp. FxanaC1)	8e-68 (93%)
SNA01168	Protoporphyrinogen oxidase (*Streptomyces auratus*)	0.0 (77%)
SNA01169	Acetyl-CoA acetyltransferase (*Streptomyces* sp. FxanaC1)	0.0 (94%)
SNA01170	Flavoprotein oxidoreductase (*Streptomyces* sp. FxanaC1)	0.0 (94%)
SNA01171	Uroporphyrinogen decarboxylase (*Streptomyces albulus*)	0.0 (91%)
SNA01172	Hypothetical protein (*Streptomyces auratus*)	2e-123 (85%)
SNA01173	LuxR family transcriptional regulator (*Streptomyces albulus* PD-1)	3e-134 (100%)
SNA01174	Hypothetical protein (*Streptomyces* sp. W007)	3e-13 (46%)
SNA01175	Hypothetical protein (*Streptomyces aurantiacus*)	1e-22 (56%)
SNA01176	3'-5' exonuclease (*Streptomyces* sp. FxanaC1)	0.0 (92%)
SNA01177	Acetyl-CoA acetyltransferase (*Streptomyces avermitilis*)	0.0 (97%)
SNA01178	3-hydroxyacyl-CoA dehydrogenase (*Streptomyces auratus*)	0.0 (95%)
SNA01179	Hypothetical protein (*Streptomyces auratus*)	4e-59 (84%)

Table A5 – BLASTp analysis of S1C5 encoding genes

Gene	Best hit description	E-value (% Identity)
SNA01188	ABC transporter substrate-binding protein (*Streptomyces* sp. Fxanac1)	0.0 (89%)
SNA01189	Apha-1 2-mannosidase (*Streptomyces* sp. FxanaC1)	0.0 (78%)
SNA01190	Hypothetical protein cf54_33035 (*Streptomyces* sp. Tu 6176)	3e-39 (57%)
SNA01191	Aconitate hydratase (*Streptomyces auratus*)	0.0 (96%)
SNA01192	Hypothetical protein (*Streptomyces auratus*)	2e-11 (48%)
SNA01193	RNA polymerase sigma 70 (*Streptomyces* sp. FxanaC1)	1e-144 (75%)
SNA01194	UDP-n-acetylglucosamine 1-carboxyvinyltransferase (*Streptomyces sp*. FxanaC1)	0.0 (96%)
SNA01195	Integrase (*Streptomyces* sp.)	0.0 (86%)
SNA01196	Endonuclease DDE (*Streptomyces albulus* PD-1)	0.0 (84%)
SNA01197	Acetyltransferase (*Streptomyces auratus*)	2e-113 (87%)
SNA01198	Hypothetical protein (*Streptomyces auratus*)	0.0 (86%)
SNA01199	ATP-binding protein (*Streptomyces auratus*)	6e-53 (73%)
SNA01200	Nudix hydrolase (*Streptomyces auratus*)	2e-64 (80%)
SNA01201	Aminoglycoside phosphotransferase (*Streptomyces auratus*)	2e-29 (73%)
SNA01202	Amidinotransferase (*Streptomyces* sp. C)	3e-145 (79%)
SNA01203	Mitogen-activated protein kinase kinase kinase kinase 5-like (*Ailuropoda melanoleuca*)	2.2 (52%)
SNA01204	Hypothetical protein (*Streptomyces scabrisporus*)	1e-74 (72%)
SNA01205	Antibiotic biosynthesis protein (*Frankia alni*)	1e-72 (62%)
SNA01207	Transposase (*kitasatospora setae*)	2e-104 (49%)
SNA01208	Sodium/hydrogen exchanger (*Micromonospora aurantiaca* ATCC 27029)	8e-100 (55%)
SNA01209	Alpha/beta hydrolase fold protein (*Catenulispora acidiphila* DSM 44928)	3e-100 (49%)
SNA01210	Chlorinating enzyme (*Catenulispora acidiphila* DSM 44928)	2e-159 (71%)
SNA01211	Putative non-ribosomal peptide synthetase (Actinokineospora sp. EG49)	8e-07 (38%)
SNA01212	Amino acid adenylation domain-containing protein (*Catenulispora acidiphila* DSM 44928)	3e-163 (53%)
SNA01213	Thioesterase (*Streptomyces viridosporus*)	4e-77 (55%)
SNA01214	Peptidase s9 (*Streptomyces purpureus*)	0.0 (64%)
SNA01215	Cytochrome p450 monooxygenase (*Nocardiopsis* sp. FU40)	9e-146 (54%)
SNA01216	Ferredoxin (*Nocardia brasiliensis* ATCC 700358)	2e-12 (64%)
SNA01217	Cytochrome P450 hydroxylase (*Streptomyces albulus*)	4e-174 (67%)
SNA01218	ABC transporter (*Streptomyces violaceusniger* Tu 4113)	5e-100 (57%)
SNA01219	ABC transporter (*Saccharomonospora cyanea*)	4e-147 (69%)
SNA01220	Hypothetical protein (*Amycolatopsis nigrescens*)	2e-82 (48%)
SNA01221	hypothetical protein (*Streptomyces natalensis*)	7e-119 (100%)
SNA01222	Acyltransferase 3 (*Frankia sp*. EAN1pec)	5e-67 (46%)
SNA01223	Antibiotic synthesis protein mbth (*Smaragdicoccus niigatensis*)	4e-35 (76%)
SNA01224	Putative l-ornithine 5-monooxygenase (*Streptomyces aurantiacus*)	0.0 (61%)
SNA01225	Putative nonribosomal peptide synthetase (*Streptomyces griseovariabilis* subsp. bandungensis)	0.0 (50%)

Table A5 – BLASTp analysis of S1C5 encoding genes (continuation)

Gene	Best hit description	E-value (% Identity)
SNA01226	Non-ribosomal peptide synthetase (*Rhodococcus hoagie*)	0.0 (43%)
SNA01227	Pyruvate carboxyltransferase (*Nocardia brasiliensis* ATCC)	3e-164 (67%)
SNA01228	Isopropylmalate isomerase large subunit (*Nocardia brasiliensis* atcc 700358)	0.0 (74%)
SNA01229	Isopropylmalate isomerase small subunit (*Nocardia brasiliensis* ATCC 700358)	9e-87 (68%)
SNA01230	Aminotransferase class I/II (*Nocardia brasiliensis*)	3e-160 (63%)
SNA01231	Cytochrome P450 (*Saccharopolyspora spinosa*)	0.0 (65%)
SNA01232	TetR family transcriptional regulator (*Streptomyces albulus*)	4e-92 (59%)
SNA01233	Hypothetical protein (*Acinetobacter baumannii*)	8.6 (44%)
SNA01234	Hypothetical protein rw1_060_00180 (*Rhodococcus wratislaviensis* NBRC 100605)	2e-87 (51%)
SNA01235	Hypothetical protein (*Streptomyces* sp. CNT372)	0.0 (59%
SNA01236	Methylated-dna- (protein)-cysteine s-methyltransferase dna binding protein (*Streptomyces* sp. S4)	0.0 (72%)
SNA01237	Putative transposase (*Streptomyces himastatinicus*)	3e-171 (74%)
SNA01238	Phage integrase (*Streptomyces scabiei* 87.22)	4e-133 (83%)
SNA01239	MerR family transcriptional regulator (*Streptomyces avermitilis*)	4e-161 (82%)
SNA01240	Transposase, partial (*Streptomyces scabiei* 87.22)	0.0 (87%)
SNA01241	Hypothetical protein (*Streptomyces sp*. MspMP-M5)	6e-111 (94%)
SNA01242	6-phosphogluconate dehydrogenase (*Streptomyces* sp. CNB091)	1e-64 (54%)
SNA01243	O-methyltransferase family protein (*Nostoc punctiforme* PCC 73102)	9e-64 (33%)
SNA01244	Biopterin-dependent aromatic amino acid hydroxylase family protein (*Gordonia sp*. NB4-1Y)	1e-57 (45%)
SNA01245	Pyridoxal-dependent amino acid decarboxylase (*Amycolatopsis nigrescens*)	0.0 (71%)
SNA01246	L-amino-acid oxidase (*Saccharothrix espanaensis* DSM 44229)	6e-57 (34%)
SNA01247	Methylase (*Streptomyces clavuligerus*)	2e-93 (50%)
SNA01248	2'-carbamoyl transferase (*Streptomyces* sp. SANK 62799)	0.0 (54%)
SNA01249	Alpha-parvin (*Chlorocebus sabaeus*)	6.8 (47%)
SNA01250	DeoR family transcriptional regulator (*Streptomyces clavuligerus* ATCC 27064)	1e-144 (80%)
SNA01251	Hydroxylase (*Streptomyces auratus*)	1e-134 (74%)
SNA01252	Acyl-coa dehydrogenase (*Streptomyces* sp. FxanaC1)	0.0 (91%)
SNA01253	AMP-dependent synthetase (*Streptomyces* sp. FxanaC1)	0.0 (85%)
SNA01254	PaaX family transcripitonal regulator (*Streptomyces* sp. FxanaC1)	2e-159 (86%)
SNA01255	Had-superfamily hydrolase (*Streptomyces auratus*)	2e-152 (92%)
SNA01256	Transcriptional regulator, TetR family (*Streptomyces ipomoeae*)	1e-18 (33%)
SNA01257	Dehydrogenase (*Streptomyces albulus*)	2e-160 (74%)
SNA01258	KsbA (TetR family) *Streptomyces auratus*)	9e-128 (89%)
SNA01259	Enoyl-coa hydratase (*Streptomyces rimosus*)	2e-158 (89%)
SNA01260	Salicylyl-coa 5-hydroxylase (*Streptomyces* sp. FxanaC1)	0.0 (85%)

Table A6 – BLASTp analysis of S5C20 encoding genes

Gene	Best hit description	E-value (% Identity)
SNA06231	PadR family transcriptional regulator (*Streptomyces albulus*)	8e-111 (82%)
SNA06232	Hypothetical protein (*Streptomyces sp. FxanaC1*)	0.0 (85%)
SNA06233	Hypothetical protein (*Streptomyces sp*. TOR3209)	1e-99 (78%)
SNA06234	Hypothetical protein (*Streptomyces hygroscopicus*)	4e-141 (91%)
SNA06235	Hypothetical protein (*Longispora albida*)	2e-23 (43%)
SNA06236	Hypothetical protein (*Streptomyces bingchenggensis*)	1e-98 (56%)
SNA06237	Hypothetical protein (*Streptomyces himastatinicus*)	0.0 (73%)
SNA06239	GntR family transcriptional regulator (*Streptosporangium roseum* DSM 43021)	3e-146 (69%)
SNA06240	LysR family transcriptional regulator (Thermobifida fusca YX)	4e-59 (45%)
SNA06241	Aminotransferase (*Streptomyces sp*. CNT302)	0.0 (76%)
SNA06242	Lysine 2,3-aminomutase (*Streptomyces sp*. AA4)	0.0 (65%)
SNA06243	Transcriptional regulator, AsnC family (*Kutzneria sp*. 744)	5e-48 (55%)
SNA06244	Integrase (*Streptomyces auratus*)	1e-10 (38%)
SNA06245	AfsR family transcriptional regulator (*Streptomyces griseus*)	1e-170 (53%)
SNA06246	LuxR family transcriptional regulator (*Streptosporangium roseum*)	4e-38 (44%)
SNA06247	Hypothetical protein (*Amycolatopsis nigrescens*)	2e-43 (63%)
SNA06248	Hypothetical protein (*Streptomyces flavochromogenes*)	3e-15 (58%)
SNA06249	Hypothetical protein (*Nocardiopsis chromatogenes*)	1e-49 (50%)
SNA06250	3-oxoacyl- (acyl-carrier-protein) synthase III (*Bacillus thuringiensis* serovar finitimus YBT-020)	1e-80 (31%)
SNA06251	Hypothetical protein (*Nocardiopsis prasina*)	0.073 (28%)
SNA06252	Hypothetical protein (*Nocardiopsis chromatogenes*)	2e-97 (30%)
SNA06253	3-ketoacyl-ACP reductase (*Clostridiales*)	1e-20 (23%)
SNA06254	Malonyl-CoA-[acyl-carrier-protein] transacylase (*Desmospora sp*. 8437)	3e-23 (78%)
SNA06255	Malonyl CoA-ACP transacylase (*Wolbachia pipientis*)	0.009 (24%)
SNA06256	Glycine amidinotransferase (*Streptomyces sp*. CNB091)	0.0 (45%)
SNA06257	Mixed NRPS PKS (*Streptomyces sp*. K01-0509)	0.0 (53%)
SNA06258	Thioesterase (*Mycobacterium*)	1e-54 (37%)
SNA06259	Hypothetical protein (*Salinispora pacifica*)	2e-19 (40%)
SNA06260	Flavoprotein (*Patulibacter americanus*)	2e-27 (37%)
SNA06261	Serine/threonine protein kinase (*Kutzneria albida* DSM 43870)	4e-122 (49%)
SNA06262	Hypothetical protein (*Salinispora arenicola*)	2e-33 (40%)
SNA06264	Hypothetical protein DOTSEDRAFT_164870 (*Dothistroma septosporum* NZE10)	6.3 (49%)
SNA06265	Hypothetical protein (*Streptomyces sp.* FxanaC1)	1e-71 (78%)
SNA06268	Deoxycytidine triphosphate deaminase (*Streptomyces albulus*)	2e-130 (95%)
SNA06269	Phosphoribosyltransferase (*Streptomyces sp*. MspMP-M5)	2e-110 (96%)
SNA06270	Ricin B lectin (*Streptomyces sp*. FxanaC1)	0.0 (86%)
SNA06271	Membrane protein (*Streptomyces sp*. MspMP-M5)	4e-149 (77%)
SNA06272	Ribonuclease HI (*Streptomyces sp*. FxanaC1)	1e-125 (81%)
SNA06273	Drug:H+ antiporter-2 (14 Spanner) (DHA2) family drug resistance MFS transporter (*Streptomyces sp*. HGB0020)	0.0 (75%)
SNA06274	TetR family transcriptional regulator (*Streptomyces sp*. AW19M42)	2e-47 (52%)

Table A6 – BLASTp analysis of S5C20 encoding genes (continuation)

Gene	Best hit description	E-value (% Identity)
SNA06275	Ribonuclease HI (Streptomyces sp. FxanaC1)	1e-125 (81%)
SNA06276	Fe-S oxidoreductase (*Streptomyces sp*. FxanaC1)	0.0 (94%)
SNA06277	Molecular chaperone DnaK (*Streptomyces griseus* subsp. *griseus* NBRC 13350)	0.0 (92%)
SNA06278	Acetyltransferase GCN5 (*Streptomyces auratus*)	1e-64 (78%)
SNA06279	Drug:H+ antiporter-2 (14 Spanner) (DHA2) family drug resistance MFS transporter (*Streptomyces sp*. HGB0020)	0.0 (75%)
SNA06280	Molecular chaperone DnaK (*Streptomyces sp*. AW19M42)	0.0 (92%)

Table A7 – Typical roles of the different families of regulators identified in the selected clusters

Regulator	Role (Reference)	Genes (cluster)
AbaA-family	Unknown (Fernández-Moreno *et al.*, 1992)	*SNA00403* (S1C2)
AfsR-family	Positive (Horinouchi *et al.*, 1990)	*SNA06245* (S5C20)
AraC-family	Positive (Soisson *et al.*, 1997, Martin and Rosner, 2001)	*SNA00399* (S1C2) *SNA00433* (S1C2)
AsnC-family	Negative (Pérez-Rueda and Collado-Vides, 2001)	*SNA06243* (S5C20)
CRP/FNR-family	Positive (Kolb *et al.*, 1993, Korner *et al.*, 2003)	*SNA00096* (S1C1)
DeoR-family	Negative (Skerlová *et al.*, 2014)	*SNA01250* (S1C5)
GntR-family	Negative (Hillerich and Westpheling, 2006)	*SNA06239* (S5C20)
LuxR-family	Positive (Chen and Xie, 2011)	*SNA01173* (S1C4) *SNA06246* (S5C20)
LysR family	Positive (Schell, 1993)	*SNA00088* (S1C1) *SNA06240* (S5C20)
MerR-family	Positive (Brown *et al.*, 2003)	*SNA01239* (S1C5)
PaaX family	Negative (Niewerth *et al.*, 2013)	*SNA01254* (S1C5)
PadR-family	Negative (Gury *et al.*, 2004, Fibriansah *et al.*, 2012)	*SNA06231* (S5C20)
Regulator	Unknown	*SNA00428* (S1C2) *SNA00830* (S1C3) *SNA06248* (S5C20)
RNA polymerase sigma factor	Positive (Gruber and Gross, 2003)	*SNA01193* (S1C5)
Serine/threonine protein kinase	Unknown	*SNA06261* (S5C20)
TetR-family	Negative (Ramos *et al.*, 2005)	*SNA00851* (S1C3) *SNA01232* (S1C5) *SNA01256* (S1C5) *SNA01258* (S1C5) *SNA06274* (S5C20)
XRE-family	Negative (Durante-Rodriguez *et al.*, 2010)	*SNA00405* (S1C2) *SNA00835* (S1C3)

7. Appendix II

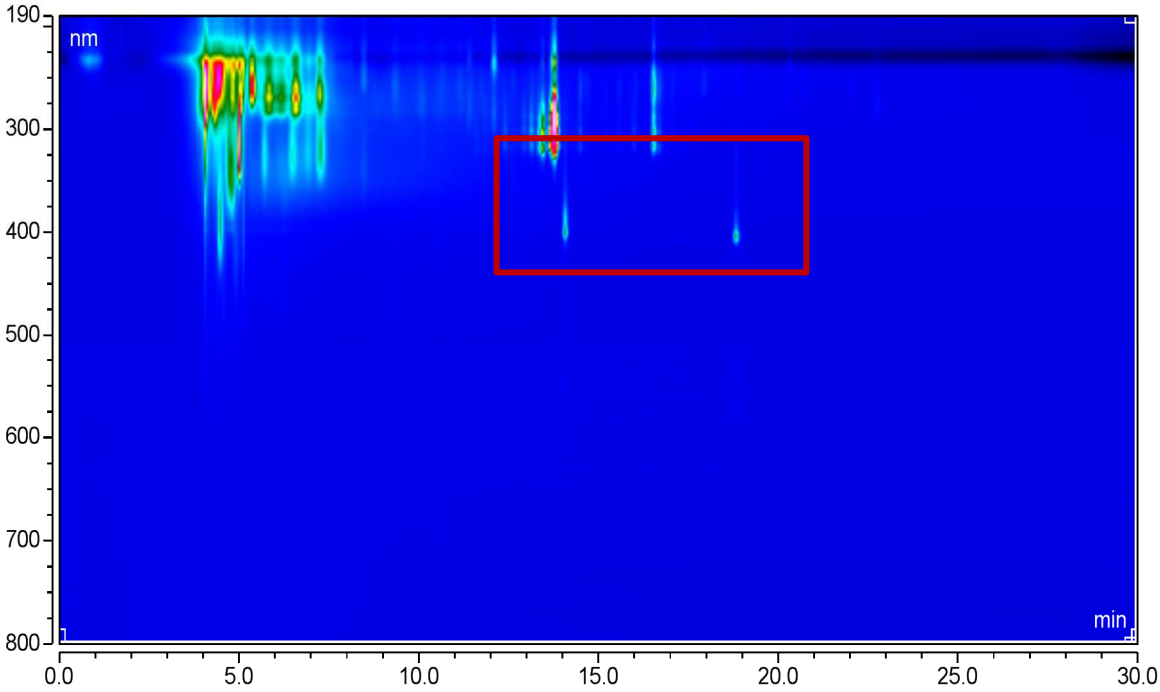

Figure A1 – HPLC diode array analysis of *S. natalenis* wild-type culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.

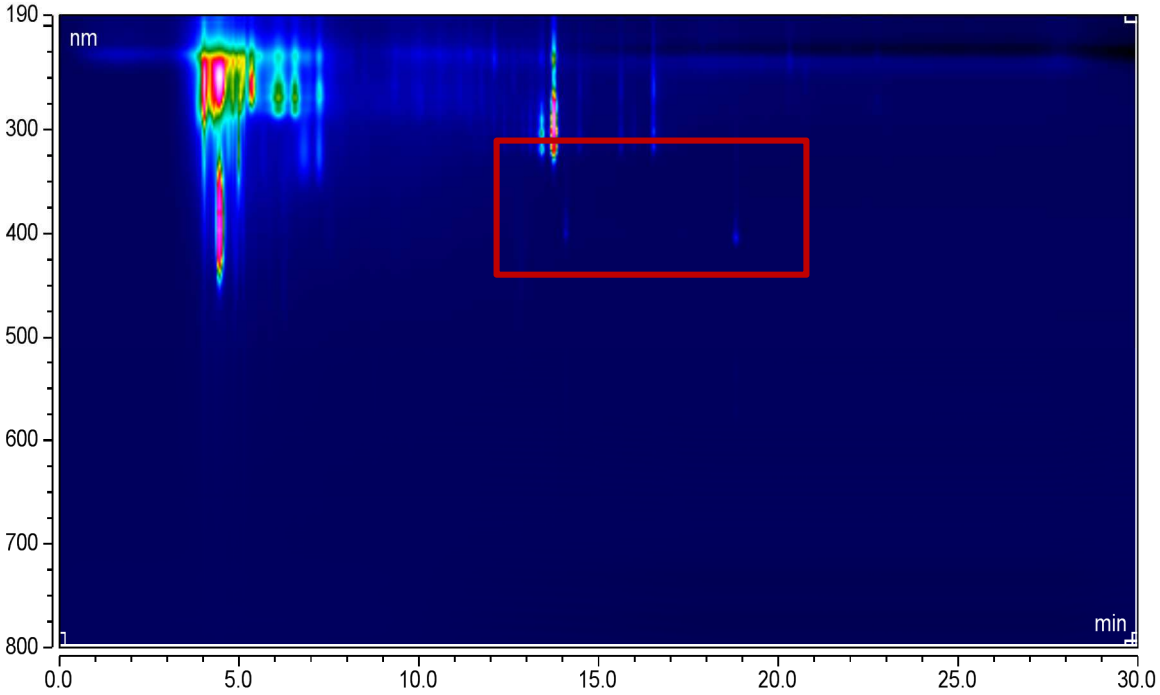

Figure A2 – HPLC diode array analysis of *S. natalenis* [pIB139] culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.

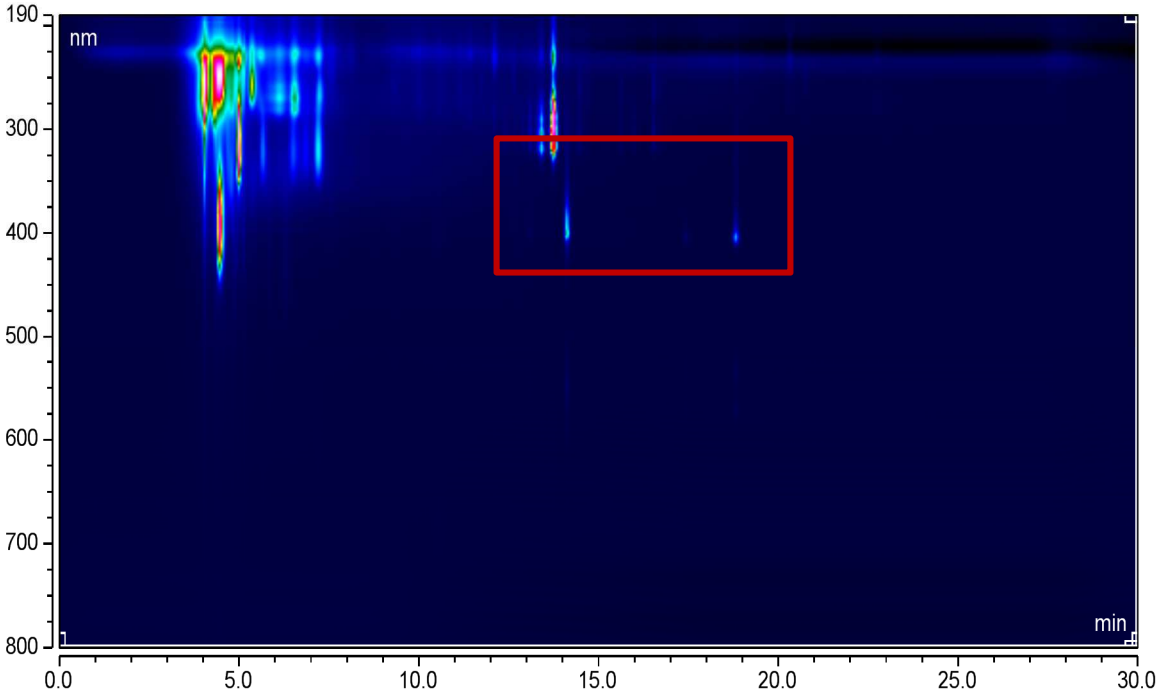

Figure A3 – HPLC diode array analysis of *S. natalenis* [pIBSNA06246] culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.

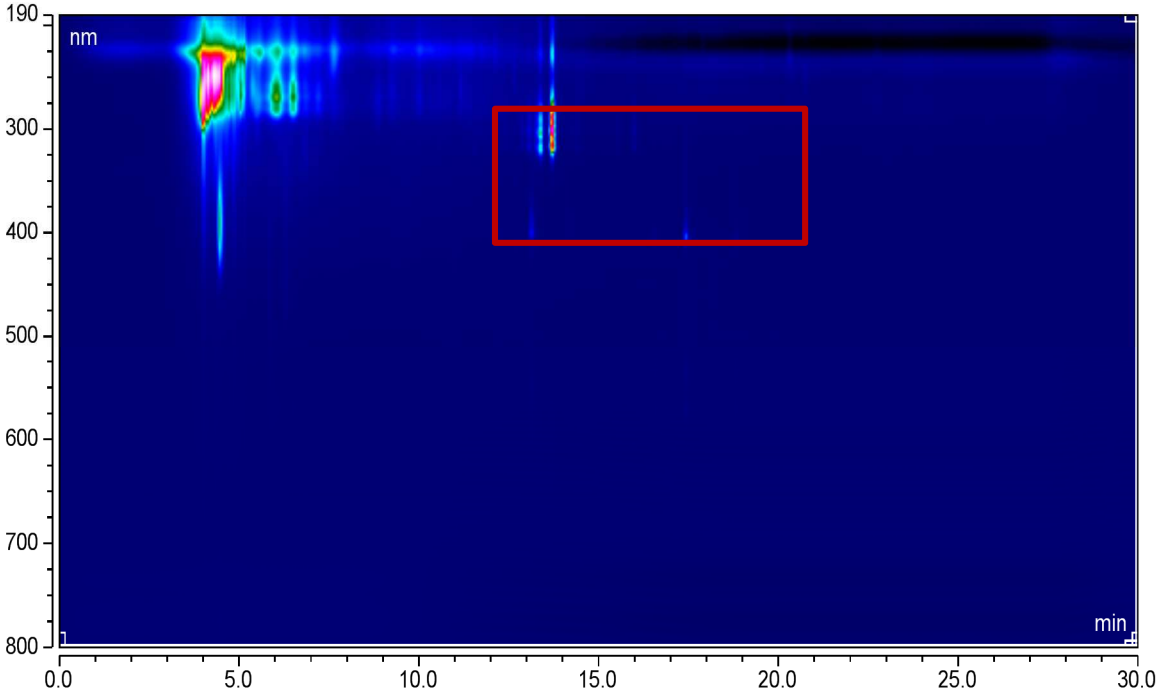

Figure A4 – HPLC diode array analysis of *S. natalenis* [pIBSNA01239] culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.

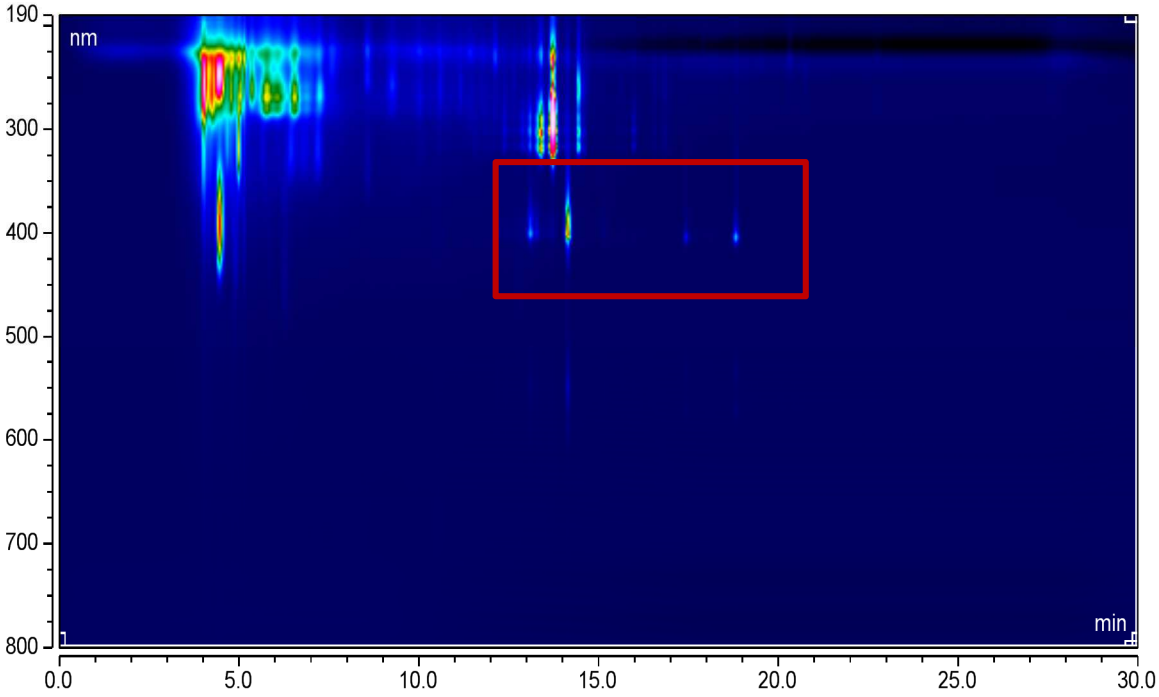

Figure A5 – HPLC diode array analysis of *S. natalenis* [pIBSNA01173] culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.

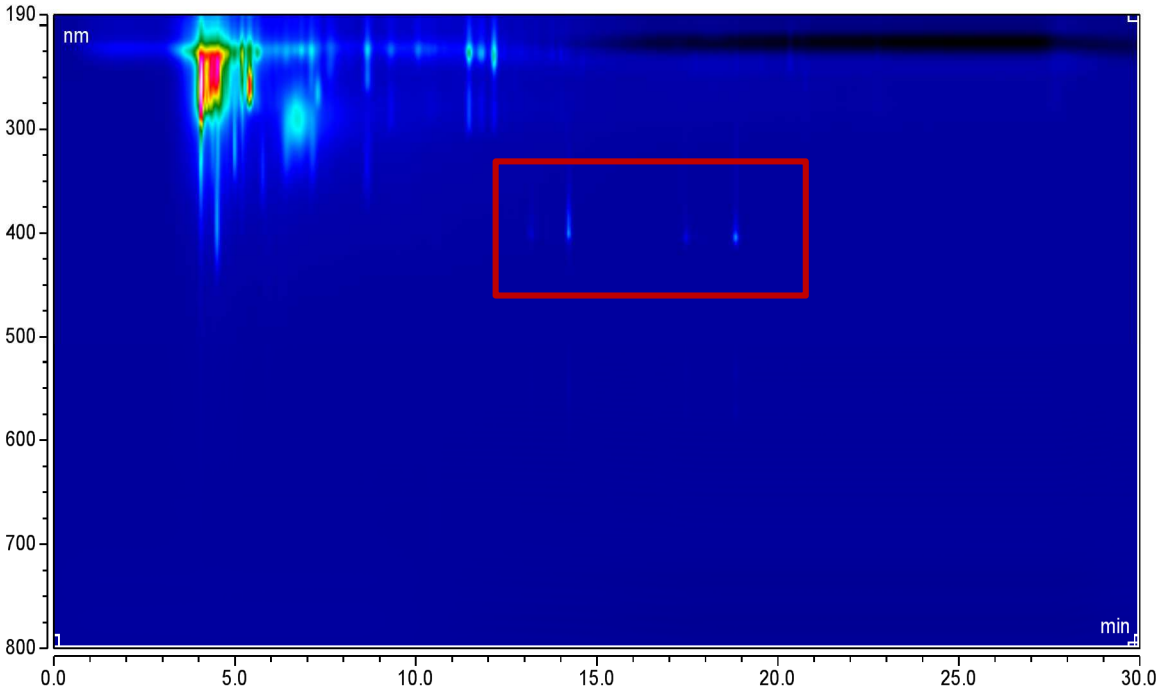

Figure A6 – HPLC diode array analysis of *S. natalenis* Δ*pimM* culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.

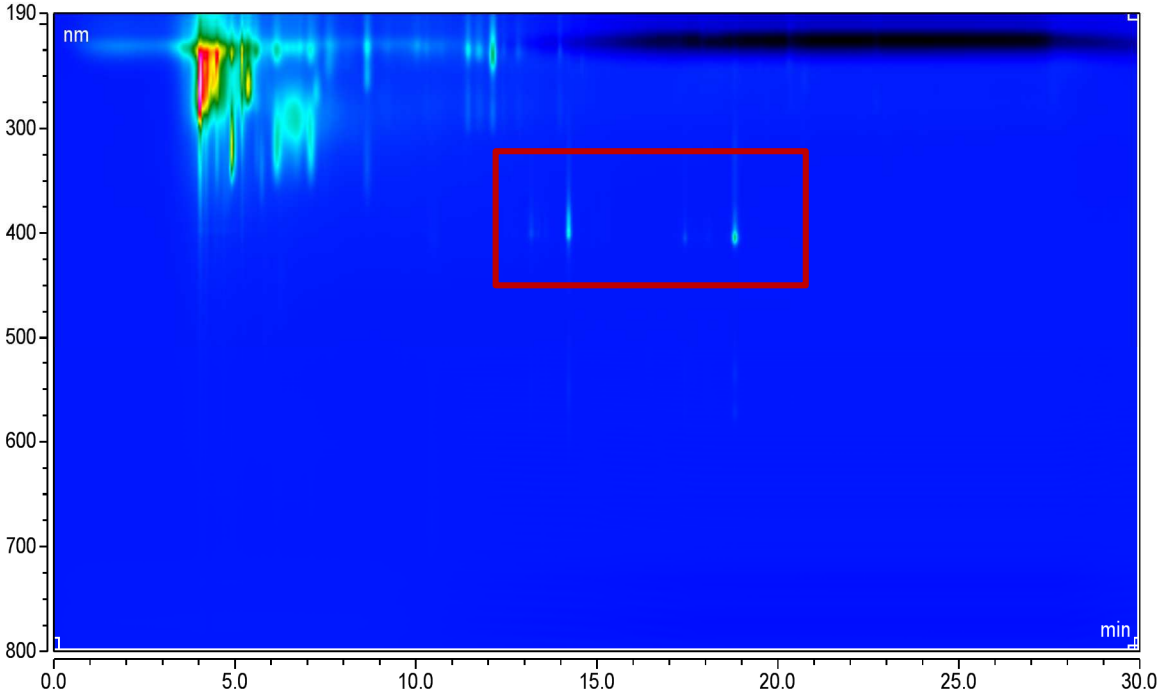

Figure A7 – HPLC diode array analysis of *S. natalenis* Δ*pimM* [pIB139] culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.

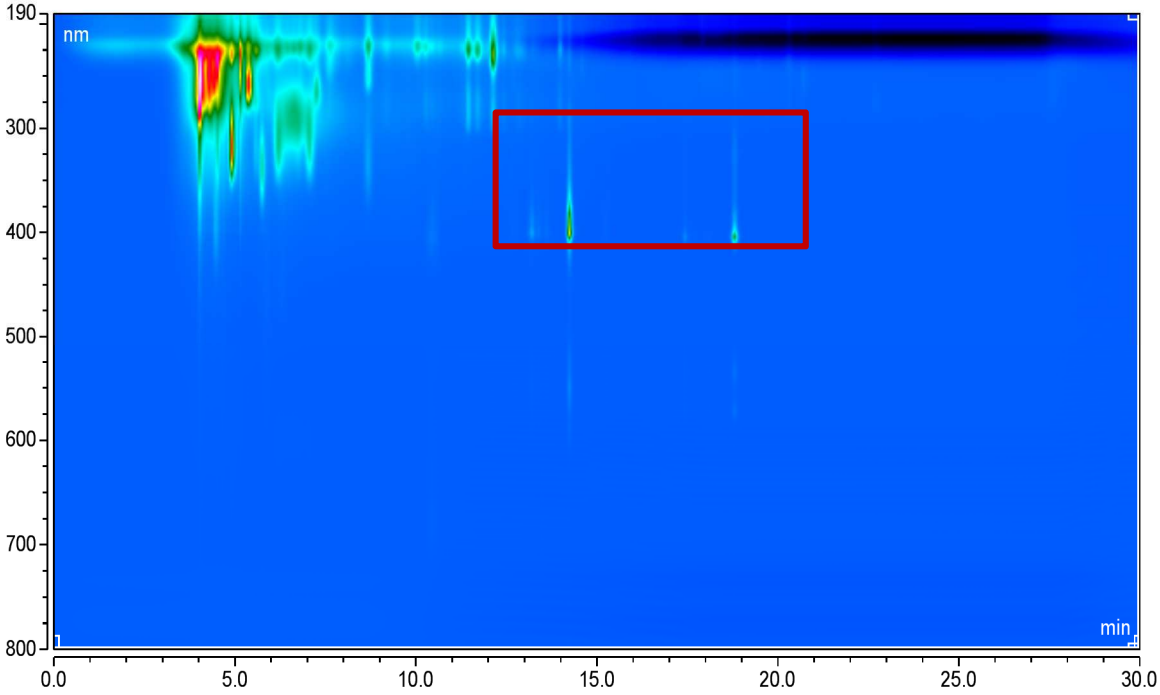

Figure A8 – HPLC diode array analysis of *S. natalenis* Δ*pimM* [pIBSNA06246] culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.

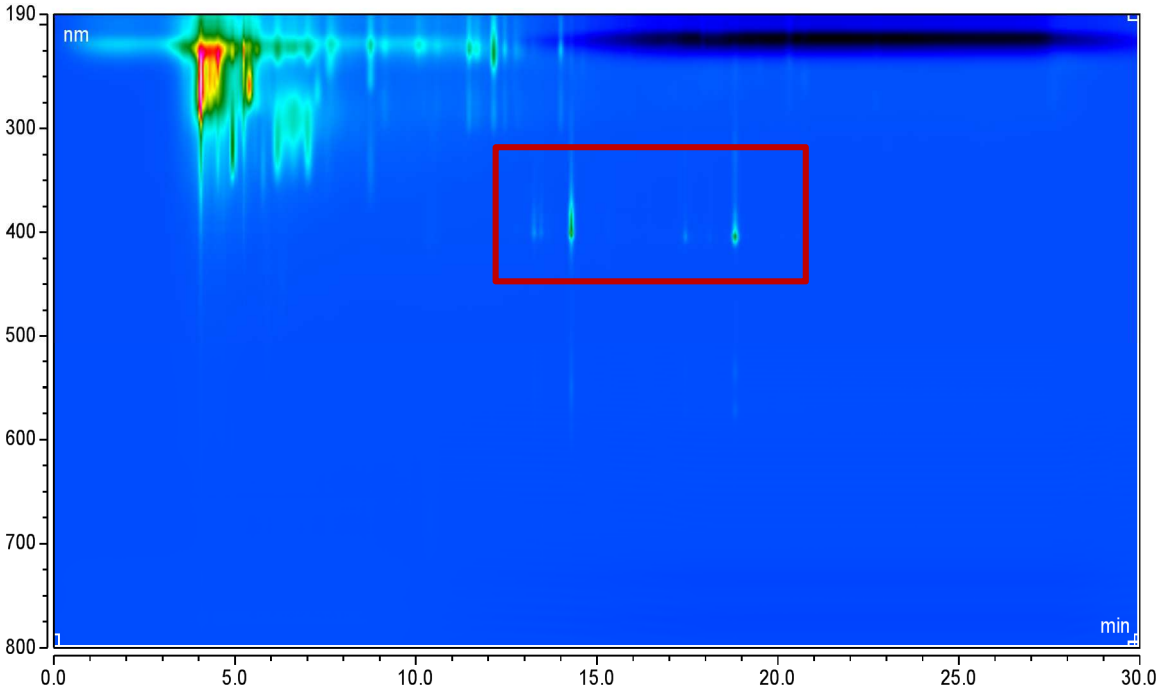

Figure A9 – HPLC diode array analysis of *S. natalenis* Δ*pimM* [pIBSNA01239] culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.

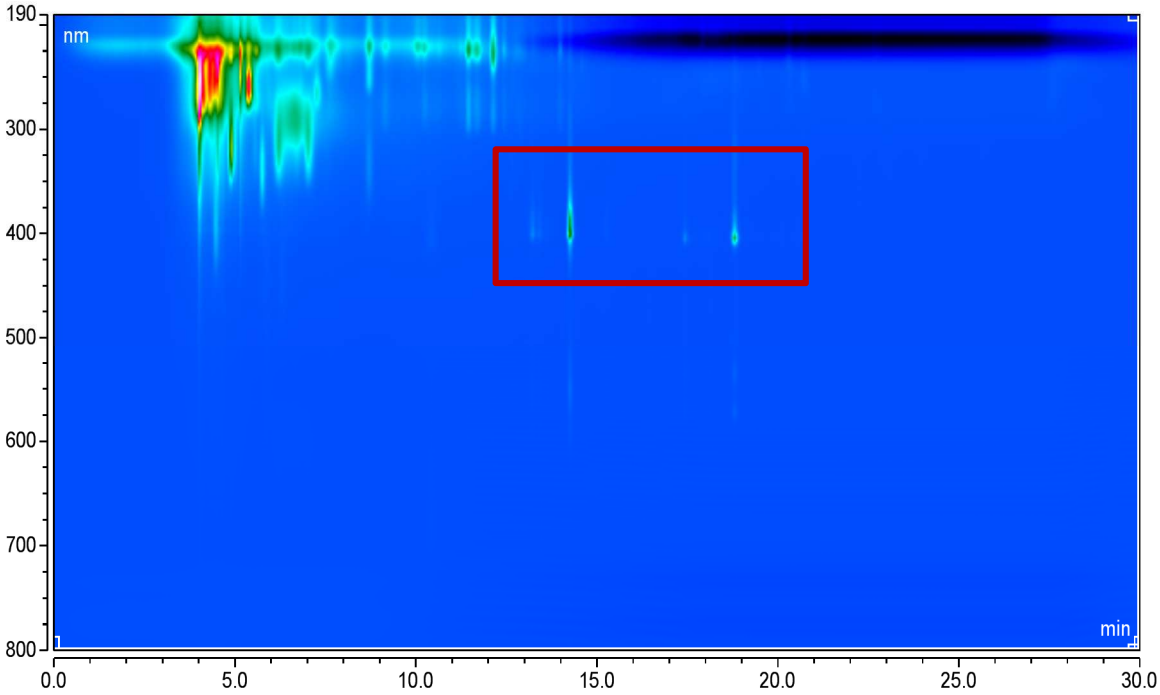

Figure A10 – HPLC diode array analysis of *S. natalenis* Δ*pimM* [pIBSNA01173] culture broth extract. The wavelength is expressed (nm) as a function of retention time (min). The red box marks the area that varies the most among all strains extracts.